Universidad Autónoma de Madrid
Departamento de Física de la Materia Condensada

Signal generation and detection in a cellular context

Tesis doctoral presentada por
F. Javier Estrada Díez

Programa de doctorado de Física de la Materia Condensada

Director:
Raúl Guantes Navacerrada

Madrid, agosto de 2012

Contents

List of Figures

List of Tables

List of acronyms

This is a list of the acronyms used in the text:

- **ODE** Ordinary differential equation
- **LNA** Linear Noise Approximation
- **FS** Feedback (feedforward) strength
- **LC** Linear cascade
- **FB** Feedback
- **P-FB** Positive feedback
- **N-FB** Negative feedback
- **FFL** Feedforward loop
- **C-FFL** Coherent feedforward loop
- **I-FFL** Incoherent feedforward loop
- **FDT** Fluctuation-Dissipation Theorem
- **AM** Amplitude modulated
- **FM** Frequency modulated
- **BW** Bandwidth
- **BW**$_{fluc}$ Fluctuations bandwidth
- **BW**$_{osc}$ Oscillations bandwidth
- **SC** Spearman's rank correlation coefficient
- **MIC** Mutual information coefficient
- **PCA** Principal components analysis

Publications

Part of the work presented in this thesis has given rise to the following publications:

- "*Dynamic and structural constraints in signal propagation by regulatory networks*"

  Javier Estrada and Raúl Guantes, (submitted 2012).

- "*Proteus: a web-based, context-specific modelling tool for molecular networks*"

  Florian Gnad,Javier Estrada and Jeremy Gunawardena, Bioinformatics **28**, 1284–1286 (2012).

- "*Trade-offs and noise tolerance in signal detection by genetic circuits*"

  Raúl Guantes, Javier Estrada and Juan F. Poyatos, PLoS ONE **5**, e12314 (2010).

Abstract

This thesis takes a Systems Biology approach to propose new methods for the study of the molecular interactions underlying fundamental cellular processes. From signal transduction to gene expression, the strongly nonlinear nature of most of these systems make them perfect candidates to be subjected to detailed mathematical modeling, in an effort to find the elements (species, interactions or degradations) in those networks which allow them to exhibit the responses which help keeping the whole organism alive. Thus, our analysis is based on that approach, and allows us to reveal the interactions present in a specific signal transduction pathway as well as to characterize the effect that the structure of simple modules of biochemical interactions has on their ability to process and propagate signals.

After a brief introduction to the cellular context where all these processes occur, we have started describing the results obtained in the study of the IP_3-mediated calcium signaling pathway. This pathway presents calcium oscillations upon stimulation of the cell with constant concentration of diverse ligands. Moreover, it seems to somehow multiplex the information encoded in the ligand type and concentration in the temporal pattern exhibited by those oscillations. By means of experiments performed in Hela cells using microfluidic devices and computer analysis of previously published models describing this system, we have been able to unveil the internal structure of the network of biochemical interactions which gives rise to such a complex behavior. All this process has been performed taking into account cell-to-cell variation and avoiding direct intervention in the molecular machinery of the network, thus avoiding the uncertainty caused by those techniques. In addition, the obtained description of the molecular machinery has allowed us to make predictions for the expected cellular responses under specific conditions. This has led us to propose and perform some experiments which have further corroborated our model.

In the last chapter of the thesis, we have described the tools developed to study the effect that different interactions cause in the dynamical and steady state response of simple gene networks, and the results obtained when applying them. First, by taking advantage of the modular description of gene networks, we have proposed a simple three component module as a general platform to test the influence that specific network interactions (feedback, feedforwards or autoregulations) have on the response of the module to amplitude and frequency modulated signals, as well as on its ability to deal with the random fluctuations inherent to gene expression. We have found that some network structures (feedbacks and autoregulations) exhibit trade-offs in the detection of both classes of signals, while other are

able to overcome that constraint, being capable of improving the propagation of both types of stimuli. At the same time, we have found that different types of circuits deal with noise differently: while some structures have to increase their signal to noise ratio to perform a feasible transmission of noisy signals, others are able to filter noise in the frequency domain by separating the range of frequencies where oscillations are best propagated from that of the fluctuations. Finally, using the simple three component network as a model, we have explored the connections between the structure of the networks, their signal propagation abilities, and their response to sudden changes in the input concentration. A statistical analysis has allowed us to obtain answers which do not depend on the specific regime the circuits are operating in, giving us useful information about the effect that different interactions have on the steady state and dynamic response of the circuits, as well as on the connection between this dynamical behavior and the steady state response.

Resumen

Esta tesis toma un enfoque típico de la Biología de Sistemas para proponer nuevos métodos útiles en el estudio de las interacciones moleculares responsables de muchos procesos celulares. La naturaleza no lineal de la mayoría de estos sistemas, desde la transducción de señal hasta la expresión genética, los convierte en perfectos candidatos para ser modelizados matemáticamente, en un intento por encontrar los elementos básicos que dan lugar a los comportamientos que permiten a la célula seguir con vida. Nuestro análisis toma precisamente este enfoque, permitiéndonos dar tanto una descripción detallada de las interacciones moleculares presentes en un camino de transducción de señal, como del efecto que la estructura de módulos sencillos de interacciones bioquímicas tiene en la capacidad de éstos para procesar y transmitir señales.

Después de dar una breve introducción sobre el contexto celular en el que ocurren todos estos procesos, hemos empezado describiendo los resultados obtenidos en el estudio de la vía de señalización de calcio mediado por IP_3. Esta vía da lugar a oscilaciones de calcio intracelular cuando la célula es estimulada con diferentes ligandos a concentración constante. Los patrones mostrados por dichas oscilaciones dan la impresión de multiplexar la información contenida en el tipo de ligando y su concentración. Por medio de experimentos hechos con células Hela dispuestas en dispositivos microfluídicos, y analizando modelos matemáticos propuestos en estudios previos, nuestro método nos ha permitido dilucidar la estructura interna de la red de interacciones moleculares que da lugar a un comportamiento complejo como este. Todo el proceso se ha llevado a cabo teniendo en cuenta la variabilidad celular y sin intervenir directamente en la maquinaria interna de la célula, evitando así la posible alteración incontrolada del sistema estudiado cuando se aplican estas técnicas. Así mismo, las descripción del sistema que hemos obtenido nos ha permitido hacer predicciones sobre la respuesta de la célula ante diferentes condiciones, lo que nos ha llevado a proponer algunos experimentos que ayudan a corroborar nuestro modelo.

En el último capítulo de la tesis hemos descrito, por un lado, las herramientas desarrolladas para el estudio del efecto que tienen diferentes interacciones en el comportamiento dinámico y en el equilibrio de redes de genes sencillas y, por otro, los resultados obtenidos con dichas herramientas. Para empezar, hemos hecho uso de la descripción modular de las redes de genes para proponer un módulo sencillo de tres componentes como modelo general en el que comprobar la influencia que tienen algunas interacciones (feedbacks, feedforwards o autorregulaciones) en la respuesta del módulo a señales moduladas en amplitud y en fre-

cuencia, así como en su capacidad para gestionar el ruido inherente a la expresión genética. Lo que hemos encontrado es que algunas estructuras (feedbacks) presentan trade-offs entre la detección de señales moduladas en amplitud y en frecuencia, mientras que otras (feedforwards) son capaces de superar esa limitación, incrementando su capacidad para transmitir ambos tipos de señal. Así mismo, hemos visto que no todos los circuitos gestionan el ruido de la misma manera: mientras que algunos sólo pueden aumentar su relación señal/ruido para transmitir la información fielmente, otros son capaces de filtrar el ruido en el espacio de frecuencias al transmitir las señales oscilatorias en un rango distinto del típico en el que fluctúa el ruido. Para terminar, siempre haciendo uso del módulo de tres componentes, hemos hecho un análisis de la relación entre la estructura del circuito, su capacidad para propagar señales y su respuesta a cambios bruscos en la concentración del estímulo. Por medio de un análisis estadístico hemos sido capaces de obtener resultados que no dependen del régimen en el que estén operando los circuitos, lo que nos ha permitido extraer conclusiones generales sobre la influencia que cada interacción tiene en la respuesta en el equilibrio y en el comportamiento dinámico del sistema, así como de la relación entre ese comportamiento dinámico y la respuesta en el equilibrio.

1. Introduction

In this first chapter we will give a short description of the context in which this thesis takes place: we describe the main mechanisms used by cells to sense their environment, and which enable them to adapt to changing situations by triggering the proper responses. In addition, we provide a first introduction to the modeling methods we apply to study some of these processes.

1.1. Stimulus detection and response

From bacteria to multicellular organisms, cells have to adapt to their environment responding to different external stimuli. Depending on the nature of these stimuli, a broad range of different signals arrives to them, causing an equally vast selection of responses: during chemotaxis bacteria sense nutrients and other chemicals which ultimately activate the intracellular mechanisms necessary to swim twoards the food source or to avoid a danger [1]. In multicellular organisms, the Notch signaling pathway is activated when cells make physical contact. This pathway is strongly involved in development as well as many other processes [2]. Growth factors secreted by the surrounding cells stimulate cell proliferation when binding to specific membrane receptors [3]. Pulsed secretion of the human growth hormone by the pituitary gland stimulates growth and cell regeneration [4].

While the molecular mechanisms involved in signal detection and propagation strongly depend on the nature of the stimulus and the subsequent reaction, the process follows a common scheme: a signal arrives to the cell, starting a series of molecular reactions which ultimately trigger specific responses. The whole process, starting with signal detection and finishing with the cell response is called signal transduction. A specific example of a signal transduction pathway will be studied in chapter 2, where the detailed structure off the IP_3-mediated calcium pathway is unveiled.

Although the cellular responses triggered by signal transduction pathways lie at different levels into the cell machinery, almost all pathways end up affecting gene expression, the deepest cellular-control stage [5]. In chapter 3 we will show how gene networks have their own signal detection and propagation mechanisms, which regulate gene expression in response to transduced signals.

1.1.1. Signal transduction

In order to process all the signals they receive, cells have evolved a diverse set of molecular mechanisms which allow them to deal with different intensities and dynamical ranges presented by different stimuli, as well as to coordinate the subsequent responses. Perhaps the more complex signal-detection abilities are found in cells building up multicellular organisms. This cells have to live, divide and die in a coordinated fashion which keeps the whole organism alive. To perform this task, cells communicate with their neighbors as well as with the rest of the organism. Depending on the range of this communication, four different systems are defined: juxtacrine, autocrine, paracrine and endocrine.

The juxtacrine system refers to interactions in which cells are making physical contact and exchange information thanks to receptors lying in their outer membrane. This kind of cellular communication is important in both development and adult life of multicellular organisms, coordinating vital actions like cell differentiation and proliferation. Examples of this kind of interaction are the previously mentioned Notch pathway, the Hedgehog and the Wnt pathways.

The autocrine system refers to a group of signals which are secreted by a cell to trigger a reaction on itself. A good example would be the growth factor secretion by T-cells, which stimulate their own proliferation to coordinate the response to an external antigen.

With the paracrine system cells communicate with their closest neighbors by secreting signals which usually have a relatively fast degradation rate, and therefore do not reach farther locations in the body. The best example of this system would probably be the secretion of neurotransmitters by neurons within the synaptic space.

Finally, the endocrine system is used to coordinate cell responses through the whole organism. Hormones secreted by endocrine glands are distributed through the body by the bloodstream, triggering a global response. The previously mentioned secretion of human growth hormone is a good an example of this type of signal.

Types of ligands

Signaling molecules are detected by the cell by binding to specific receptors. These receptors can be expressed in the cellular membrane or inside the cytosol, depending on the ability of the ligand to diffuse through the lipid bilayer: hydrophilic molecules have to be detected by memebrane receptors, while hydrophobic ligands are capable of diffusing through the cell membrane.

Among the signaling molecules which are able to bind directly to receptors inside the cytosol or even inside the nucleus we have the family of the steroid hormones, together with thyroid hormone, vitamin D_3 and retinoic acid. Testosterone, estrogen and progesterone are steroid hormones produced by the gonads. Corticosteroids are produced by the adrenal gland and regulate functions like the glucose production or the regulation of kidney activity.

Ecdysone activates metamorphosis in insects, while the brassinosteroids regulate development in plants. Vitamin D_3, thyroid hormone and retinoic acid are not steroid hormones, but cells use the same mechanism to deal with them. The thyroid hormone is synthesized in the thyroid gland, and plays a main role in metabolism regulation. Vitamin D_3 regulates calcium metabolism and bone production in vertebrates while retinoid acid is important in development. Although these molecules belong to different families, they all intervene with the cellular machinery by interacting with steroid hormone receptors. These receptors are transcription factors whose activity is regulated by the ligands, which trigger their function as gene activators or repressors. A diagram showing the estrogen activation is provided in Fig.(1.1)

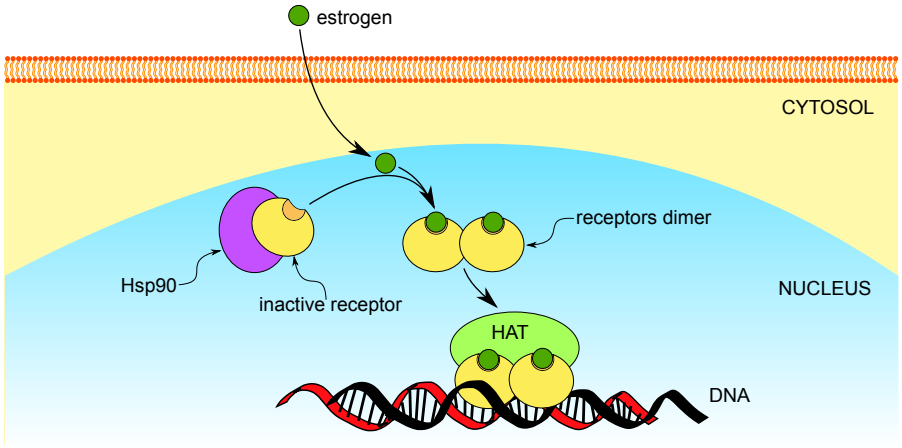


Figure 1.1.: Estrogen action: estrogen diffuses through the cellular membrane directly into the nucleus. There, it binds to the receptor which was previously bound to Hsp90. The receptor dimerizes and binds to DNA to work as a transcription factor for its target genes upon binding with co-activator HAT.

Nitric oxide (NO) and carbon monoxide (CO) are extremely simple molecules which are also able to diffuse through the cell membrane to directly activate specific molecular mechanisms. Unlike steroid hormones, these molecules do not bind to transcription factors. Instead, they interact with specific enzymes which subsequently trigger a specific response. NO has a mean lifetime of the order of seconds, and therefore it only gets to the closest cells. It is involved in vasodilation, by propagating a response originally initiated by a neurotransmitter through the cells in the surrounding tissue.

Neurotransmitters are another type of signal which coordinate responses in the whole organism. They are secreted by neurons, and trigger a reaction in neighbor neurons or other cells, like, for instance, muscle cells. Among these molecules we find dopamine, serotonin or histamine. Some of these molecules may work not only as neurotransmitters, but as hormones as well. This is the case of epinephrine, which works as a neurotransmitter but is also secreted by the adrenal gland to activate glycolysis in muscle cells. Neurotransmitters are

small hydrophilic molecules which are unable to directly diffuse through the cellular membrane. Thus, the signals they carry are detected by the cell by binding to receptors expressed in the cellular membrane. Among these receptors there are ligand-triggered ion channels, which allow the passage of specific ions into the cytosol upon binding to the neurotransmitter, and G protein-coupled receptors, which are trans-membrane proteins which suffer an structural transformation when they bind to the ligand, leading to a subsequent cascade of reactions inside the cell.

Peptides are the most abundant signaling molecules in animals. Among this group we find peptide hormones like glucagon, together with hormones produced in the pituitary gland like the growth hormone. Neuropeptides are secreted by some neurons to work as neurotransmitters or neurohormones, which interact with neurons farther apart. Nerve growth factor (NGF) and epidermal growth factor (EGF) belong to the polypeptide group. NGF's regulate neuron development and EGF stimulated cell proliferation. An example of the action of a growth factor would be the response caused by the secretion of platelet-derived growth factor (PDGF). PDGF is secreted during coagulation, and stimulates cell proliferation to start the wound healing process. All these molecules are unable to diffuse through the lipid bilayer, and therefore interact with specific receptors expressed in the cellular membrane.

Eicosanoids are lipids which rapidly hydrolyze, and are therefore used for autocrine and parecrine signaling. Among them we find prostacyclin, thromboxan and leukotriene. They are involved in platelet aggregation, inflammation and contraction of smooth muscle tissue. Again, this molecules are detected by the cell by binding to specific membrane receptors.

Types of membrane receptors

G protein-coupled receptors are the most abundant receptors in the cell membrane. Upon binding of a ligand to its external part, the receptor suffers a change in its cytosolic configuration, activating and releasing G proteins. Then, G proteins transmit the external signal by working as enzymes or by activating a ionic channel Fig.(1.2). The human genome expresses many different types of G proteins which are associated with specific receptors, and therefore link this receptors with specific signaling pathways. This kind of receptor is the first component in the IP_3 mediated pathway that will be analyzed in chapter 2.

Unlike G protein-coupled receptors, receptors tyrosine kinase directly bind to intracellular enzymes. Among this family we find receptors for the polypeptide growth factors. The human genome codes for 58 receptor of this kind, which detect ligands like EGF, NGF, PDFG or insulin.

There are other types of membrane receptors, like cytokine receptors, receptors tyrosine phosphatase or receptors for the transforming growth factor β family, which plays a main

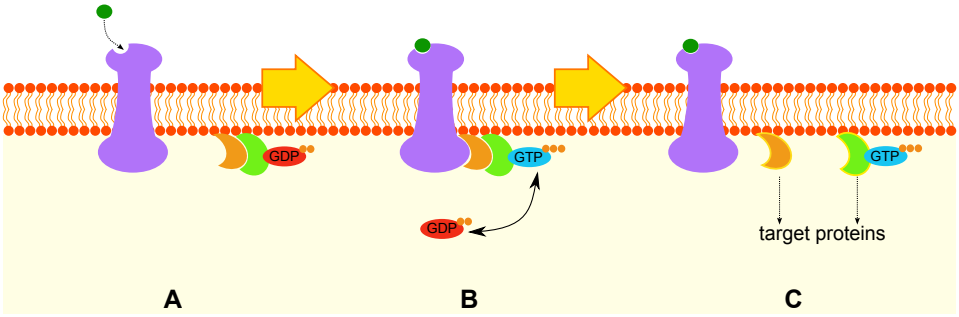


Figure 1.2.: G protein-coupled receptor: while in its inactive state (A), the α subunit (light green) of the G protein binds to GDP, building up a complex together with the β and γ subunits (orange). Upon binding of a hormone (dark green) to the receptor (purple), this interacts with the G protein, stimulating the interchange of GDP and GTP (B). The α subunit binds to GTP and the $\beta\gamma$ complex separates and interacts with its corresponding target proteins to continue the signal transduction process (C).

role in inhibiting cell growth, inflammation and collagen metabolism.

Second messengers

As it has been already described, the typical signal transduction mechanism consists on ligands which interact with receptors in the cell membrane. These receptors ultimately activate specific enzymes, known as second messengers, which start a series of reactions inside the cell which propagate, amplify and process the signal, leading to a specific response.

An example of the action of a second messenger can be found in the cyclic adenosine monophosphate (cAMP) pathway. This pathway is involved in glycogen-glucose transformation in muscle cells: upon activation of a specific G protein due to an increase in epinephrine concentration, cAMP concentration increases. cAMP works as a second messenger by transporting the original signal carried by the hormone -the first messenger-, through the cell to activate a series of reactions which ultimately end up catalyzing the transformation of glycogen into glucose. This pathway is involved in the more complex voltage-induced calcium pathway, which would be explained in detail in chapter 2. Another example of the second messenger action will be discussed in much detail in chapter 2, where the role of inositol (1,4,5) trisphosphate (IP_3) in calcium dynamics is studied in detail.

So far we have described how signals are received and propagated inside the cell. Some of the responses triggered by these signals remain in a superficial layer, with all the decision making process taking place at the signal transduction level. In general, this happens with decisions which need to be taken fast, like the ones involved in chemotaxis: in bacteria, Che proteins are phosphorylated and de-phosphorylated upon interaction with signals triggered by membrane receptors, which sense nutrients in the extracellular media. These proteins regulate the tumbling frequency and the membrane receptor action itself, playing a main

role in motor behavior in response to attractant or repellent stimuli [6]. Nevertheless, most of the signals end up interacting with transcription factors which ultimately regulate gene expression, therefore affecting the cell fate at the deepest level.

1.1.2. Gene expression

The deepest stage of the cell decision making machinery lies at the gene expression level. Although all the information needed to build a living organism is encoded in its genome, what ends up arising from that pool of information is the specific expression state of the genome: which genes are active and which ones silenced at a given time. This expression state plays a main role in both long and short term processes during the organism life: on the one hand, it regulates the different stages of the organism cycle, when to divide or differentiate, when to arrest the cycle or when to initiate apoptosis. On the other hand, most signal transduction pathways ultimately affect gene expression in order to respond to signals which regulate faster decision-making processes, like flagellum formation in bacteria [7] or regulation of the cell metabolic rate in vertebrate cells [8].

When a gene is active, it is accessible to the molecular machinery involved in the process by which the information it encodes is synthesized into an effective product. In general, this product is a protein, although there are non-protein coding genes which encode sequences of functional RNA's such as transfer RNA's or ribosomal RNA's. Proteins work as main effectors in most cellular processes, they build up functional structures like receptors or cilia and flagellum, and play a main role in regulating the expression state of other genes. The process by which the information encoded in an active gene ends up forming a protein consists of three steps: first transcription takes place, RNA polymerase binds to the promoter of the gene, a specific region which indicates the beginning of the coding sequence, and starts reading the sequence of nucleotides while creating an analogous RNA copy called messenger RNA (mRNA). Second, once the whole sequence has been transcribed, mRNA exits the nucleus into the cytoplasm, where it binds with ribosomes which start translating its nucleotide sequence into an amino-acid chain. To finish, this amino-acid chain folds mainly due to the interactions between its different regions to give rise to a protein in its functional from.

Protein syntheses, and therefore, the expression of protein-coding genes, is regulated at different levels: on the one hand, there is regulation at the transcription level, where genes are activated or repressed by different transcription factors. These transcription factors can be either the result of upper-stream signal transduction reactions, or the product of other genes. On the other hand, there is also translational regulation, in which the translation of the already transcribed mRNA is regulated by other components , like micro RNA's (miRNA) or small interfering RNA's (siRNA) [9–11].

All the positive and negative regulations which control gene expression at different levels,

end up forming an extremely complex network of molecular interactions which play a main role in coordinating almost every single cellular process, and allow cells to respond and adapt to their changing environment Fig.(1.3).

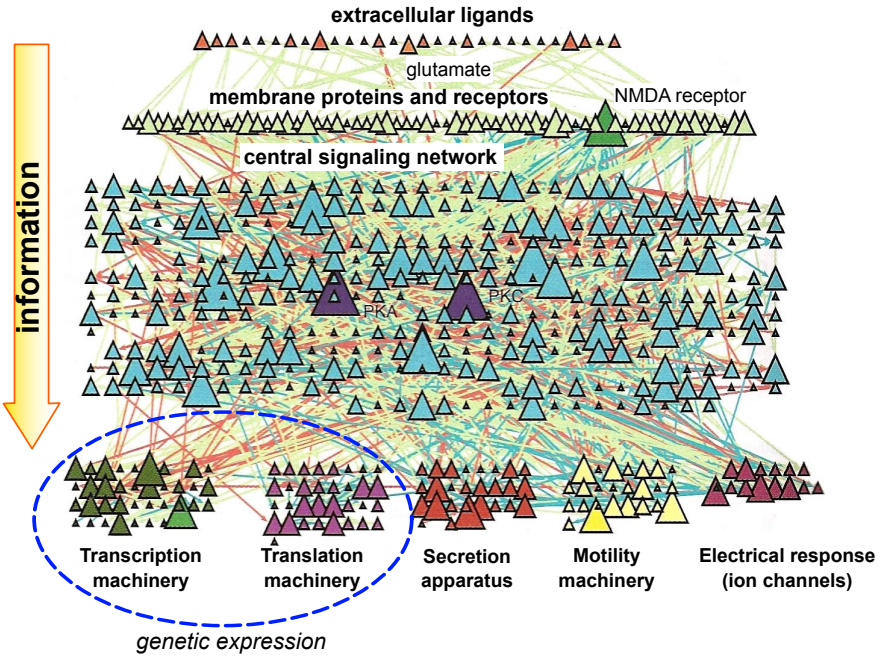


Figure 1.3.: Whole mammalian signaling pathway: green and red lines respectively state for activating and repressing interactions, while blue ones state for neutral links between elements. Extracellular ligands (red triangles) are detected by membrane receptors (light green triangles), which activate specific signal transduction pathways (blue triangles). The central signaling network interacts with the gene expression machinery mainly by activating and repressing genes at the transcription level (dark green triangles). Figure extracted from [12].

Modularity

Given the complexity of the gene networks described above, it may seem hopeless trying to understand their overall behavior. Nevertheless, during the last fifteen years, detailed analysis of complex networks have shown that most of them share some statistical features which give some hints about their global organization [13, 14]. Gene networks are classified as 'scale-free' networks [14]: the fraction of nodes (genes) having k edges, $p(k)$, decays as a power law $p(k) \sim k^{-\gamma}$ (with γ around 2 or 3). This kind of networks are found in other contexts as well, like metabolic networks [15], scientific collaboration networks [16] or the World Wide Web [17].

With the motivation of going beyond this first classification, Alon and coworkers analyzed different complex networks trying to unveil simpler structures lying inside the whole structure of nodes and edges [18]. In their work, they looked for the occurrence of patterns

comprised of 3 and 4 interacting nodes and analyzed their structure. Given a specific network, they found that some of those patterns, named 'network motifs', are significantly more abundant than expected if the network structure would be random Fig.(1.4). Moreover, when conducting this analysis on gene and neuronal connectivity networks, they found that they both share most of their network motifs, while food webs or electronic circuits present very different patterns. They stated that the differences and similarities found for these networks may be related with the overall function of the network: while gene and neuronal networks process information, food webs are just predator-prey relations. Following this idea, one would expect individual network motifs found in information-processing networks to be more efficient performing signal-processing tasks. With this in mind, several works have studied specific signal-propagation properties of these network motifs, trying to unveil the evolutionary constraints that favored their selection [19–21].

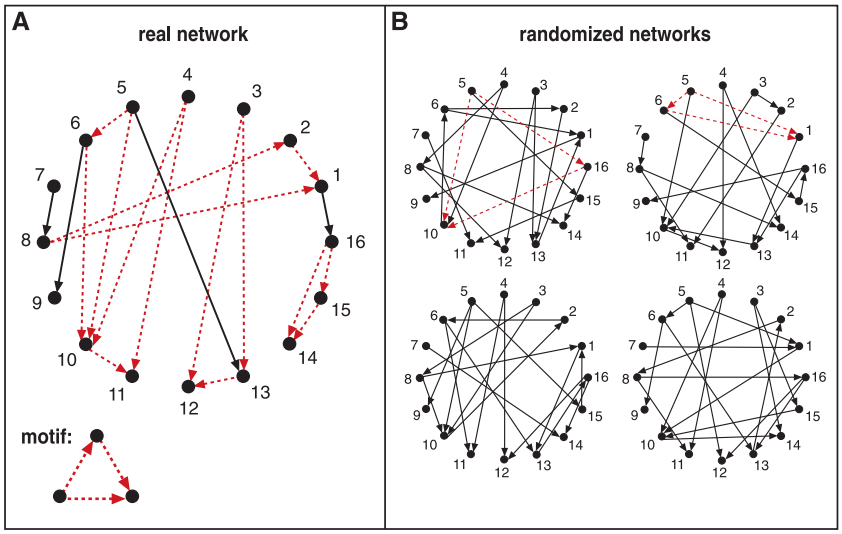


Figure 1.4.: Network motifs in a 16 node real network (A) compared to equivalent randomized networks (B): the specific pattern represented in red is much more represented in the real network than in similar randomized networks. Picture extracted from [18].

Nowadays the concept of modularity in gene networks is widely accepted, and is extensively applied to analyze and understand specific gene pathways, as well as to design synthetic networks able to perform specific tasks [22]. On chapter 3 of this thesis we will describe the results obtained from an extensive analysis of the dynamical and signal propagation properties of simple network motifs. This motifs are found in gene networks of evolutionary distant model organisms like *Escherichia Coli* or *Saccharomyces cerevisiae*, and therefore represent important examples of simple gene circuits strongly involved in the signal-response mechanisms of gene networks.

1.2. Types of signals and processing mechanisms

So far we have introduced the basic cellular machinery involved in signal detection and propagation. We have seen that extracellular stimuli are detected by receptors placed both in the surface of the cellular membrane and in the cytoplasm. Upon binding to an external ligand, receptors activate specific signaling pathways which process and propagate information and trigger the necessary cellular reaction. Most of these processed signals end up affecting the expression state of the organism genome, whose complex network analyzes the information and responds by activating and repressing specific genes. This modification of the gene expression leads to changes in the concentration of certain proteins and other gene products, which allow the organism to respond and adapt to the constantly changing environment.

From this perspective, we see that signal detection and propagation takes place at different stages: receptors and their associated signal transduction pathways have to detect a broad range of different signals, from subtle and sustained increments, to bursts in the concentration of specific molecules [23]. At the same time, gene circuits have to be able to detect signals propagated by the transduction mechanisms. As it will be shown in chapter 2, some signal transduction pathways are involved in the detection of many different classes of external stimuli. During the signal transduction process, these signals are multiplexed into limited types of molecules which subsequently transmit the information through the cell to trigger the appropriate responses. The fact that information originally encoded in many types of molecules is concentrated in just one, two or three second messengers, ends up creating very complex intra-cellular signals which ultimately have to be detected and interpreted by the genetic machinery.

Sometimes information is multiplexed and encoded in the concentration of an specific signaling molecule. This concentration regulates downstream processes by stimulating or repressing the synthesis of diverse products. The process can be continuous, where the synthesis of the products increases or decreases with the signal concentration, or it can exhibit a more complex behavior. For example, cyclin-dependent protein kinases (Cdks) regulate the production of some 'executor' proteins at different stages to trigger elaborated responses during the cell cycle [21]. On the other hand, information can be encoded in the dynamical behavior of the second messenger, rather than in its concentration at a given time. The cellular response to signals of oscillatory or pulsatile concentration usually depend on the period or pulse width of the stimulus, rather than on its amplitude. There are many examples of such mechanisms inside cells, with time scales ranging from seconds to hours and days. The IP_3 mediated calcium pathway exhibits calcium oscillations with periods which go from seconds, when cells are stimulated with ATP, to minutes when histamine is applied [24]. The p53 system, part of the genetic machinery involved in apoptosis initiation and cell proliferation control, oscillates with periods of around 7 hours when DNA damage is detected [25, 26]. On the other hand, circadian rhythms oscillate entrained with day-night transitions,

and therefore have periods of around 24 hours, affecting processes like hormone secretion [8] or body temperature [27].

The structure of the biochemical networks which deal with all these kinds of signals strongly depends on the type of stimulus they have to process, as well as on the responses they have to trigger. For instance, to be able to generate pulsatile intracellular signals, some kind of negative feedback has to be present [28]. On the contrary, when biochemical circuits are working as signal detectors, processing and propagating information to downstream genes, other structures are found. Some network topologies exhibit fast responses, while others propagate only sustained stimuli. As it will be discussed in detail in chapter 3, feed-forward loop structures respond fast to step-like stimuli, while linear cascades, like mitogen-activated protein kinase (MAPK) reactions, show responses which become slower with the length of the chain of reactions. Adaptation and fold-change detection are other important features of most most signaling circuits, allowing cells to respond to changes in he concentration of a given substance rather than to its actual level. This kind of mechanisms allow bacteria to swim along the food gradient during chemotaxis [29], or neurons to properly respond to tactile stimuli [30]. While in signal transduction pathways adaptation is in part achieved via receptor desensitization or saturation, the topology of the downstream circuit is also strongly involved in that feature, with structures commonly found in gene networks, like incoherent feed-forward or negative feedback loops, proven to respond to changes in the input stimulus [31, 32] rather than to its actual state.

1.3. Noise

In the previous sections we have provided a brief description of the main signal detection and propagation mechanisms used by cells to sense their environment. In addition, we have seen that these mechanisms have evolved to deal with diverse types of signals and control the proper cellular responses. Now it is time to emphasize that, when we described these signals, we were somehow giving them a deterministic nature: when we talked about information encoded in the concentration of a given molecule, we were assuming that this concentration was constant, that no fluctuations were present. In a similar way, frequency-encoded signals were treated as if they had perfectly determined amplitude and period. This assumptions allowed us to better explain the different information-encoding mechanisms, and sometimes give good descriptions of real biochemical reactions which shed light on real problems. Nevertheless, real biochemical signaling is more complicated than that.

In contrast to our initial description, biochemical reactions inside cells are intrinsically noisy, and therefore deterministic descriptions of signaling pathways and gene expression are always incomplete. Chemical reactions depend on reactants meeting each other to give rise to some products. Ligands depend on diffusion to get close enough to receptors on the

cellular membrane, to interact with them and initiate the subsequent signal transduction reactions. Similarly, gene activation by transcription factors depends on transcription factors reaching a specific region of the gene, while transcription is initiated only when RNA polymerase molecules diffuse close enough to interact with the gene promoter. Finally, protein translation depends on mRNA exiting the nucleus by diffusion and randomly meeting ribosomes in the cytoplasm. If the concentration of ligands, receptors, mRNA's and the rest of signaling molecules is high, the probability of each of these reactions occurring per unit time would be high too, and therefore deterministic models would give accurate descriptions. On the contrary, concentrations of signaling molecules tend to be low: some types of membrane receptors are present in numbers as low as 20 per cell [33], DNA usually has one copy of each gene, and the overall protein number in bacteria is very often lower than 100 per cell [34]. In addition, all these molecules are closely packed inside the cell together with other macromolecules, forming a fluid in which diffusion is slow, and therefore, in which the probability of two molecules meeting each other is further reduced Fig.(1.5) [35].

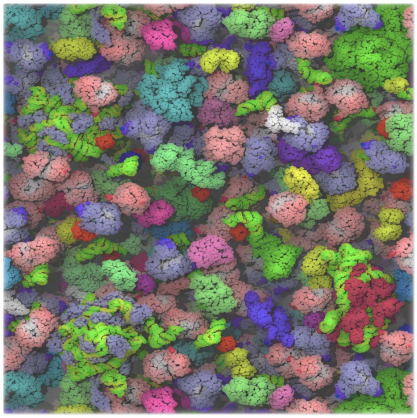


Figure 1.5.: It's crowded inside cells: molecular dynamics simulations were used to simulate diffusion inside the crowded cytoplasm of *E. Coli*. RNA is shown in green and yellow, packed together with many other molecules. Figure extracted from [36].

1.3.1. Stochastic gene expression

Stochastic gene expression is probably one of the most studied cases of noise in biochemical networks. Therefore, it is a good example to provide a more detailed description of such phenomena.

The first experimental evidence of the stochastic nature of gene expression was found in 1957 by Novick and Wiener [37], when they observed random cell-to-cell variation in the production of beta-galactosidase. Since then, randomness in gene expression has been increasingly studied, specially with the development of experimental techniques able to track gene expression at the single-cell level [38–42]. The first theoretical framework successfully

describing stochasticity in gene expression was developed in 1997-1998 by McAdams and Arkin [43, 44]. They applied Gillespie's Monte Carlo algorithm [45] to perform stochastic simulations of gene expression, giving a first explanation to the random decision making outcomes exhibited by phage lambda in the lysis-lysogeny decision making process.

As it has been already introduced, noise in genetic expression comes from the fact that gene activation-deactivation, polymerase binding, RNA transcription and protein translation are all stochastic processes. However, the final contribution to the overall expression noise of each of these reactions is different, with gene activation-deactivation producing bigger fluctuations than, for example, protein translation. Moreover, while some sources of noise equally affect most of the genome, others are independent for each gene, leading to the concept of extrinsic and intrinsic noise [46]. Since it affects the whole genome, RNA polymerase concentration represents an extrinsic noise source, while transcription initiation itself is independent in each gene, and therefore its contribution is intrinsic. Noise in gene expression has been widely studied in both prokaryotes and eukaryotes, revealing remarkable differences between both groups. While noise in prokaryotic gene expression depends mainly on the rates of transcription and translation [47], it seems to be mainly of extrinsic origin in eukariotic cells [48, 49]. Noise in eukaryotes has been usually interpreted in terms of expression bursts: periods of abundant protein synthesis, followed by periods of low activity. A detailed description of how the different components intervening in gene expression influence noise is presented in the next section, in which the main tools used to model biochemical networks are introduced.

1.3.2. The role of noise

The intrinsically noisy nature of biochemical reactions in general, and stochastic gene expression in particular, play a main role in key biological processes like evolution and adaptation. Although cells have evolved mechanisms to deal with noise when accurate information transmission is needed [50, 51], sometimes the existence of intrinsic variability in gene expression has an evolutionary advantage, allowing cells to better adapt to their surroundings. Since the environment is constantly changing, natural selection shapes organisms in a way that they can be flexible enough to deal with these variations. Therefore, although the expression level exhibited by a cell under specific conditions may not be optimal, it helps the organism to better respond to the fluctuations of the environment [52]. This randomness in gene expression allows a population of clonal cells to survive when the conditions are drastically changed: while some cells die under the new circumstances, others are able to deal with them and reproduce [53], with their expression state being inherited by the next generation. The set of gene regulation mechanisms which can be inherited by following generations is called epigenetics [54], and represents an evolutionary mechanism with two remarkable properties: on the one hand, it does not involve a modification of the genetic code itself, and, although

it is not able to generate radical changes in the organism, it is, in principle, reversible [55]. On the other hand, it is fast, with few generations needed to observe relevant changes, allowing a population to survive to fast environmental changes. Epigenetics has opened the door to the development of experimental evolution, which is helping to shed some light over fundamental problems in biology, like the evolution of multicellular organisms [56].

1.4. Modeling biochemical networks

Traditionally, Biology has been mainly a reductionist discipline: it tries to explain the characteristics of complex biological systems by understanding the individual properties of its components. On the contrary, Systems Biology takes a wider angle perspective by analyzing the interactions between these components and the emergent phenomena that subsequently arise. Although they seem somehow contradictory, both approaches are usually complementary: while descriptions of the components forming a complex system are gathered using biological and biochemical tools, Systems Biology takes advantage of these descriptions as inputs for its analysis. Within this context, it is often difficult and somehow unnecessary to make a clear distinction between Biology, Biochemistry and Systems Biology, given that ultimately they all aim to understand living organisms in a systematic way. Said so, we could say that Systems Biology does not refer to the study of specific biological problems, but to an approach taken to understand them. Its main distinctive feature is quantitative understanding of biological processes, usually combining experimental input with mathematical modeling at different levels of accuracy.

1.4.1. Deterministic models

Mathematical characterization of biological systems has been increasingly applied since the beginning of the 20th century, becoming a cornerstone in disciplines like Neuroscience and, of course, Systems Biology. In 1910 Alfred J. Lotka developed what was later known as the Lotka–Volterra model, a set of ordinary differential equations (ODE) which describes predator-prey interactions [57]. Another famous example is the the Hodgkin – Huxley model, which describes action potentials in neurons using analogies with electronic circuits [58]. The Michaelis-Menten kinetic equation is one of the simplest models describing biochemical reactions, but it allows us to introduce some concepts that will be found in almost every model used in this thesis. It was developed in 1913 by Leonor Michaelis and Maud Menten to model enzymatic reactions [59], and describes the rate at which an enzyme-mediated reaction takes place.

$$E + S \underset{k_1}{\overset{k_{-1}}{\rightleftharpoons}} ES \xrightarrow{k_2} E + P \tag{1.1}$$

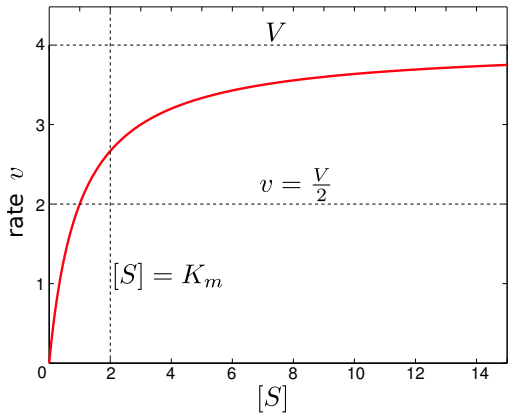


Figure 1.6.: Michaelis-Menten rate v as a function of the substrate concentration $[S]$. Following Eq.(1.2), v reaches half its maximum value V when the substrate concentration is equal to K_m.

To model this reaction, Michaelis and Menten performed a "time-scale separation". They applied a "quasi-steady state" approximation by assuming that the amount of enzime E stays constant and that the dimer ES exhibits faster dynamics than the substrate S and the product of the reaction P -and is therefore at equilibrium-. This simplification allows to express the rate of change of product concentration $[P]$ as:

$$\frac{d[P]}{dt} = \frac{V \cdot [S]}{K_m + [S]} = v \tag{1.2}$$

Here, the reaction rate v increases with the concentration of substrate $[S]$ until a maximum rate $V = k_2[E]_0$ is achieved. $[E_0]$ states for the enzyme concentration, assumed as constant. The constant $K_m = \frac{k_{-1}+k_2}{k_1}$ indicates how fast v increases: when $[S] = K_m$ the reaction rate is half its maximum value Fig.(1.6).

Time-scale separation is an important consideration to take into account when modeling biochemical reactions [60]. The quasi-steady state approximation was first introduced by Briggs and Haldane [61], and allows both to simplify the model and to test hypothesis about the characteristic time-scales of the different sub-reactions comprising the system under study. Another good example of time-scale separation can be applied to the following gene expression model described in Fig.(1.7) [62].

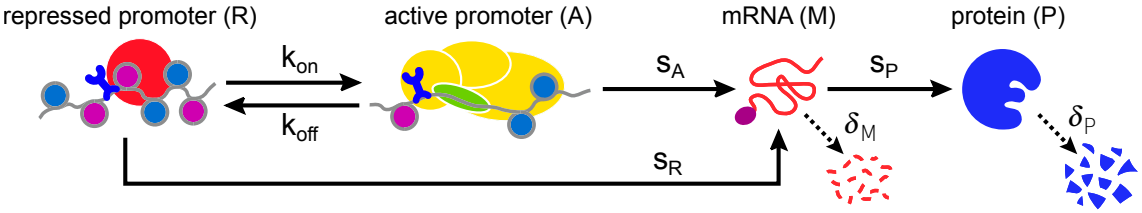


Figure 1.7.: Simple gene expression model: the gene can be either active (A) or repressed (R), with transitions between states given by k_{on} and k_{off} rates. mRNA (M) is transcribed with different rates depending on the promoter state: s_R and s_A respectively for the active and repressed states. Finally, protein (P) is translated with rate s_P, while both mRNA and protein decay with rates δ_M and δ_P respectively. Figure adapted from [62]

The gene expression model in Fig.(1.7) follows

$$\begin{aligned} \frac{d[R]}{dt} &= k_{off}[A] - k_{on}[R] \\ \frac{d[A]}{dt} &= k_{on}[R] - k_{off}[A] \\ \frac{d[M]}{dt} &= s_A[A] + s_R[R] - \delta_M[M] \\ \frac{d[P]}{dt} &= s_P[M] - \delta_P[P] \end{aligned} \tag{1.3}$$

Here, a gene can be either active (A) or repressed (R). Transitions between both states are given by k_{on} and k_{off} rates. mRNA concentration $[M]$ is transcribed with different rates depending on the gene state: the repressed state transcription rate (s_R) is 10-fold lower than the active state one (s_A). Finally, protein concentration $[P]$ is translated with rate s_P, while both mRNA and protein decay with rates δ_M and δ_P respectively.

By assuming that transcription factor binding-unbinding to a promoter occurs much faster than mRNA transcription and protein translation, we can simplify the set of equations by assuming the number of effective active and repressed promoters to be at equilibrium, A_{eff} and R_{eff}

$$\begin{aligned} \frac{d[R]}{dt} &= k_{off}[A_{eff}] - k_{on}[R_{eff}] = 0 \\ \frac{d[A]}{dt} &= k_{on}[R_{eff}] - k_{off}[A_{eff}] = 0 \end{aligned} \tag{1.4}$$

Assuming that there is only one copy of the gene ($R + A = 1$) and doing some algebra we get to

$$\begin{aligned} [R_{eff}] &= \frac{k_{off}}{k_{on} + k_{off}} \\ [A_{eff}] &= \frac{k_{on}}{k_{on} + k_{off}} \end{aligned} \tag{1.5}$$

This leads to a simplified system of equations where the original 4 species have been reduced to only 2

$$\begin{aligned} \frac{d[M]}{dt} &= s_A \frac{k_{on}}{k_{on}+k_{off}} + s_R \frac{k_{off}}{k_{on}+k_{off}} - \delta_M[M] \\ \frac{d[P]}{dt} &= s_P[M] - \delta_P[P] \end{aligned} \tag{1.6}$$

A table of characteristic time-scales in gene expression in *E. Coli* is provided in Table (1.1), showing how reasonable this approximation seems to be.

Binding of active transcription factor to its promoter	$\sim$ 1 s
mRNA transcription + Protein translation	$\sim$ 5 min
50% change in protein abundance upon gene activation	$\sim$ 1 hour

Table 1.1.: Characteristic gene expression time-scales. Adapted from [63].

Kinetic rate equations like Michaelis-Menten equation describe the behavior of individual reactions which are usually embedded in more complex systems. These complex systems are formed by different species interacting with each other: signaling cascades where proteins are phosphorylated and dephosphorylated by upstream products or networks of interacting genes. The complexity of these interactions gives rise to a broad set of spatial and temporal phenomena. To model the behavior of these bigger structures, simple component descriptions are merged into sets of differential equations. The resulting ODE systems account for the interactions between species and their dynamical behavior. Although the spatial behavior is sometimes taken into account [64], it is usually ignored by supposing that intracellular containers (cytosol, mitochondria, vesicles...) are small [65].

The kind of curve given by the Michaelis-Menten equation is recurrently found in biological systems. Biochemical reactions are in one way or another mediated by enzymes, which in this context can be anything from ribosomes to membrane receptors. Therefore, the synthesis rate of macromolecules depends on substrate concentration and saturates to a maximum value due to the limited amount of enzyme. For example, cell stimulation by an external ligand is limited by the number of membrane receptors in the cell surface, once all the available receptors are bounded to a ligand molecule, an increase in ligand concentration is not detected by the cell anymore. Another example of sigmoidal behavior is found in networks of interacting genes like the ones described in Figs (1.3 1.4). In these structures, production of a gene transcription factor is regulated by transcription factors of upstream genes. The production probability exhibits a sigmoidal behavior, and is usually modeled using Hill functions [20, 63, 66]. In general, Hill functions are useful to describe binding of a macromolecule to a ligand [67], and account for both activating (F_{act}) and repressing (F_{rep}) interactions

$$\begin{aligned} F_{act} &= \beta \frac{(u/k)^H}{(1+(u/k)^H)} \\ F_{rep} &= \beta \frac{1}{(1+(u/k)^H)} \end{aligned} \tag{1.7}$$

Here, u states for the concentration of the upstream molecule which enhances or represses the production rate. Again, k indicates at which point the Hill function shows half its maximum value. H is known as the Hill coefficient, and controls the steepness of the function. It is usually around 2 in gene expression models Fig.(1.8). β sets the maximum reaction rate.

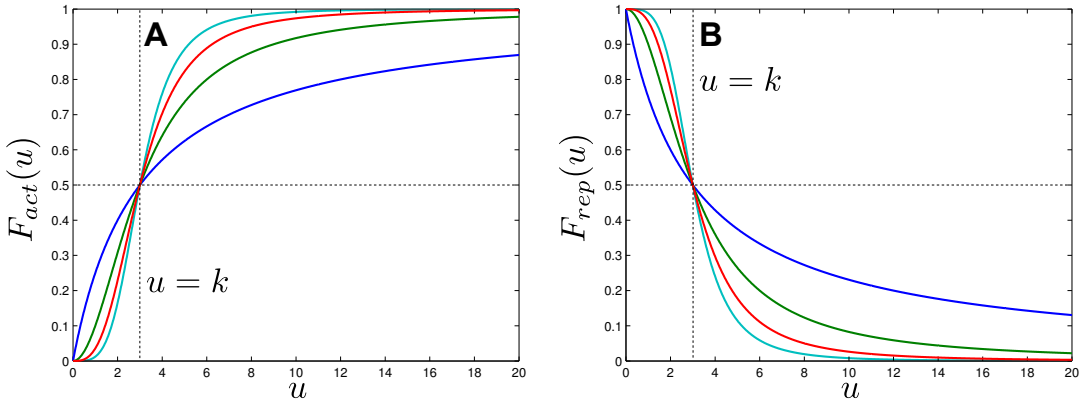


Figure 1.8.: Hill functions: activating (A) and repressing (B) equations as a function of the regulating species u, for $\beta = 1$. Half the maximum value is achieved for $u = k$. Hill coefficients $H = 1, 2, 3, 4$ are indicated in blue, green, red and cyan respectively. The steepness of the function increases with H.

In chapter 3, where we work with models of interacting genes, gene-gene interaction is modeled without taking into account any of the gene expression reactions. On the contrary, the process is simply described by activating or repressing Hill functions, depending on the sign of the interaction between both genes.

At this point it becomes clear that one of the main goals when modeling biological processes is based on simplification: we want the simplest possible model able to reproduce the behavior of our system. On the one hand, this allows us to build a first hypothesis on the most important components giving rise to the experimentally observed behavior. On the other hand, the simpler the model, the easier it would be to solve and analyze it, allowing for a deeper understanding of the relations between the involved species and their interactions, and the characteristics of the modeled phenomena: concentrations, time-scales or noise propagation, to name some examples.

Nevertheless, despite all the possible simplifications, the resulting ODE systems tend to be strongly nonlinear. This nonlinearity is in general unavoidable if one wants to reproduce complex responses, like concentration oscillations [68], generation of concentration pulses [69] or even cell differentiation and morphogenesis [70]. The drawback for this is the fact

that, in general, only numerical solutions are available, rendering detailed analyses difficult. Nevertheless, as it will be shown in detail in chapter 3, in most cases linear approximations of strongly nonlinear systems are a powerful and reliable tool to analyze trade-offs and correlations in such complicated systems.

In addition to nonlinearity, there is always an uncertainty when choosing the values for the constants in the models. Although both *in vitro* and *in vivo* experiments provide relatively good estimations of protein degradation rates, binding constants and other relevant quantities, the numerical models we deal with are almost always an extremely condensed version of the real systems, and many underlying reactions are simplified in a single parameter (for example, by making quasi steady-state dynamics assumptions). Therefore, many of the constants present in the model represent just effective degradation rates or effective binding constants, which may easily depend on different factors like cell cycle stage and specific gene expression level. This adds further complications to the analysis, although, as it will be shown in chapters 2 and 3, it is still possible to extract from the models parameter-independent properties which allow for a reliable comparison with the real system.

1.4.2. Stochastic models

Deterministic models give a first approximation to the behavior of arrays of biochemical reactions. However, due to the intrinsic noisy nature of intra-cellular processes, at some point deterministic models fail to describe some responses observed at the single-cell level, as well as cell-to-cell variation in monoclonal populations. We have already mentioned that this variability plays an important role in key biological processes like genetic selection, evolution or differentiation. Therefore, developing models to describe the stochasticity of intra-cellular reactions becomes unavoidable if a complete description of the cellular behavior has to be achieved.

In the last 15 years [43], statistical mechanics tools that had been traditionally used to model stochasticity in chemical reactions have been increasingly applied in the description of biochemical processes. In most cases the dynamics of biological systems are well described as Markov jump processes, in which any change in the state of the system occurs discretely after a random time period, with change and time both depending only on the previous state [65, 71]. This kind of description has proven to accurately account for the experimentally observed variability in different contexts [72, 73].

To provide an illustrative example, the genetic expression model introduced in the previous section Fig.(1.7) is analyzed following such stochastic description [34]. Assuming that there are ($[A_{max}] = [A] + [R]$) copies of the gene, and that there is no mRNA transcription when the gene is repressed ($s_R = 0$), we can re-write equation Eq.(1.3) as follows

$$\begin{aligned}
\frac{d[A]}{dt} &= k_{on}([A_{max}] - [A]) - k_{off}[A] = k_{on}[A_{max}] - \delta_A[A] \\
\frac{d[M]}{dt} &= s_A[A] - \delta_M[M] \\
\frac{d[P]}{dt} &= s_P[M] - \delta_P[P]
\end{aligned} \tag{1.8}$$

Here the parameter $\delta_A = k_{on} + k_{off}$ provides a characteristic time-scale for changes in gene state.

The ODE system Eq.(1.8) defines a three-variable Markov process, which can be modeled using the Master Equation $\frac{df(A,M,P)}{dt}$ [74]

$$\begin{aligned}
\frac{df(A,M,P)}{dt} &= k_{on}(A_{max} - A + 1) \cdot f(A-1, M, P) \\
&- k_{on}(A_{max} - A) \cdot f(A, M, P) \\
&+ k_{off}(A+1) \cdot f(A+1, M, P) - k_{off} A \cdot f(A, M, P) \\
&+ s_A A \cdot f(A, M-1, P) - s_A A \cdot f(A, M, P) \\
&+ \delta_M(M+1) \cdot f(A, M+1, P) - \delta_M M \cdot f(A, M, P) \\
&+ s_P M \cdot f(A, M, P-1) - s_P M \cdot f(A, M, P) \\
&+ \delta_P(P+1) \cdot f(A, M, P+1) - \delta_P \cdot f(A, M, P)
\end{aligned} \tag{1.9}$$

Each positive term of the above equation states for the probability of the system getting from adjacent states to (A, M, P). For example, $s_A A \cdot f(A, M-1, P)$ indicates the probability of getting from the $(A, M-1, P)$ state to (A, M, P) by transcribing one copy of mRNA. The rate of this reaction is proportional to the number of active genes A and the rate constant s_A. Similarly, negative terms indicate the probability of the system evolving from (A, M, P) to a neighbor state: by degradation of one of the species -for example, mRNA degradation $\delta_M M \cdot f(A, M, P) \Rightarrow (A, M, P) \rightarrow (A, M-1, P)$-, or by production of another -again, mRNA $s_A A \cdot f(A, M, P) \Rightarrow (A, M, P) \rightarrow (A, M+1, P)$-.

Being a linear equation, Eq.(1.9) could be solved exactly following the work by David R. Rigney [75, 76]. Nevertheless, we follow a different approach: as it will be described in chapter 3, solving a Lyapunov equation which depends on the Jacobian matrix of the ODE system Eq.(1.8) and a diffusion term, we can obtain a solution for the covariance matrix, and therefore, for the noise coefficient of variation in protein expression evaluated at the stationary state -defined as the standard deviation of the fluctuations divided by the mean $\sigma_P^2 / \langle P \rangle^2$-

$$\frac{\sigma_P^2}{\langle P\rangle^2} = \overbrace{\frac{1}{\langle P\rangle}}^{\text{Protein intrinsic noise}} + \overbrace{\frac{1}{\langle M\rangle}\frac{\delta_M}{\delta_M+\delta_P}}^{\text{Propagated mRNA noise}} + \overbrace{\frac{1-A_{eff}}{\langle A\rangle}\frac{\delta_M}{\delta_M+\delta_P}\frac{\delta_P}{\delta_A+\delta_P}\frac{\delta_A+\delta_M+\delta_P}{\delta_A+\delta_M}}^{\text{Propagated gene activation-repression noise}} \tag{1.10}$$

The overall noise in protein expression can be divided in three distinct terms:

The first one accounts for the intrinsic fluctuations of the protein translation process, which follows a Poisson distribution with the expected coefficient of variation of $\langle P\rangle^{-1}$.

The second term represents the contribution to the fluctuations in protein production caused by propagation of fluctuations in random mRNA transcription. Again, mRNA transcription itself is a Poisson process -$\langle M\rangle^{-1}$-, but the amplitude of its fluctuations is attenuated through time-averaging when they propagate along the network.

Finally, the third term accounts for the influence of random gene activation-repression in protein production. Being $A_{eff} = \frac{k_{on}}{k_{on}+k_{off}}$, as defined in Eq.(1.6), gene activation-repression follows a less noisy binomial distribution as long as $k_{on} >> k_{off}$. On the contrary, if the gene repression rate is much larger than the activation ($k_{on} << k_{off}$), then $1 - A_{eff} \simeq 1$, leading to a Poisson process. In this situation, the influence of fluctuations in gene activation-repression becomes an important contribution on the overall fluctuations in protein expression. Similarly to mRNA, fluctuations are attenuated during propagation along the network: from gene activation-repression to mRNA transcription and from mRNA to protein translation.

The previous case provides a good example in which exact solutions can be found for a linear system. Nevertheless, many interesting systems are strongly nonlinear, forcing us to use different approaches, both analytical and computational. Among the analytical approaches, in this thesis we extensively use the linear noise approximation (LNA), in which the Jacobian matrix of the non-linear ODE system is linearized around the equilibrium state before obtaining the covariance matrix. On the other hand, computational approaches like Gillespie's Monte Carlo algorithm [45] or Langevin approximations are the most commonly used.

Gillespie's algorithm follows these basic steps:

1) Given the state of the system and the set of reactions, the propensities of each reaction are computed.

2) In the basis of the current propensities, the time step to the next reaction is randomly generated (the larger the propensities, the shorter the time step).

3) The reaction occurring at this time step is randomly chosen based on its propensity (the larger the propensity, the more likely the reaction, following a Poisson distribution)

4) Update the system state according to which reactions have occurred and go back to step one.

Although Gillespie simulations provide a more precise solution, Langevin approximations employ shorter computing times which allow to realistically simulate even larger networks. They are based on finding a diffusion process that matches the dynamics of the Markov jump. This process is described by a stochastic differential equation which can be integrated with similar techniques as the deterministic models [77].

Gillespie's algorithm and Langevin equations are the usual methods to simulate stochastic processes. In fact, in chapter 3 of this thesis Gillespie's algorithm is extensively applied to simulate noise in gene expression and compare it with theoretical results obtained through LNA. Nevertheless, the development of new stochastic simulation algorithms is an active area of research [78–80].

As an example, Gillespie simulations are conducted following the simple model described in in Fig.(1.7), helping us to dissect the effect of the various noise sources in the system Fig.(1.9).

Noise in protein production depends on the different rate constants. Given relatively fast transitions between active and repressed states, high numbers of expressed mRNA lead to low-amplitude fluctuations in protein production Fig.(1.9-A). If mRNA production is decreased, the amplitude of the fluctuations in protein abundance increases due to the finite-number effect of low copy numbers of mRNA Fig.(1.9-B). Therefore, it is proportional to the number of proteins synthesized per RNA copy, and thus proportional to s_P/δ_M. This translational bursting mechanism seems to be responsible for noise in gene expression in prokaryotes [81]. A different mechanism of noise, called transcriptional bursting, appears when transitions between active and repressed gene states take longer to occur (given that protein decay is fast enough [34]). In this case, protein production is divided in periods of low and high abundance, in what has been called transcriptional bursting Fig.(1.9-C). In eukaryotes, this seems to be a main source of noise in protein abundance.

1.5. The evolutionary perspective

Both form and function of every biological entity is ultimately shaped by the subtle principles of natural selection. Although this is already an obvious statement, our human brain tends to think about every possible or existing device in an engineering way, being biology no exception to this rule. Since this thesis deals with the structure and behavior of different biological networks in a clear analogy with electronic circuits and similar engineered devices, one easily forgets about the big picture of the problem, in which every molecule and every interaction between molecules exist just because in the past generations a similar molecule or interaction lived for long enough to be inherited by the next generation.

Therefore, although along this thesis the reader may find lots of engineering analogies, these are just used to gain a clearer understanding of the details of the problem or to better

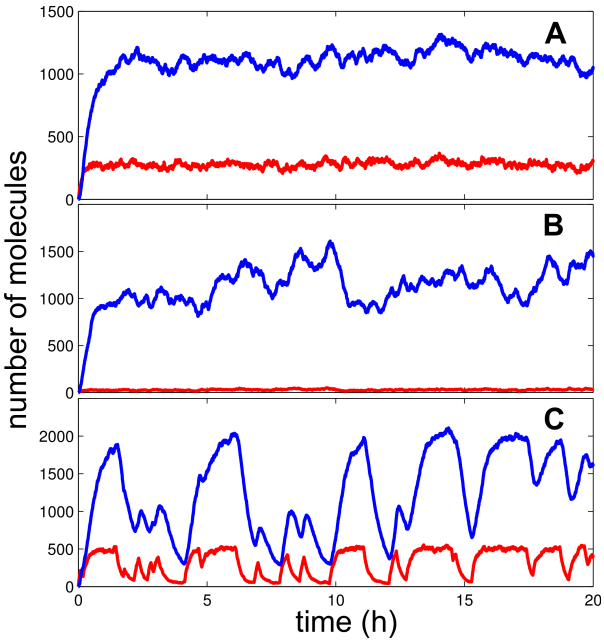


Figure 1.9.: Translational and transcriptional bursting: as a result of Gillespie simulations of model 1.3, protein (blue) and mRNA (red) abundances are plotted against time for different rates. Protein production rates are chosen in order to get always averages in protein numbers of around 1000 ($s_P = 0.2\ min^{-1}$ in A and C, and $s_P = 2\ min^{-1}$ in B). Decay rates are fixed $\delta_M = 0.1$ and $\delta_P = 0.05$ per minute. A: given fast transitions between active and repressed gene states ($k_{on} = k_{off} = 10$ per minute), low-amplitude protein fluctuations are found when mRNA is abundant ($s_A = 50\ min^{-1}$ and $s_R = 5\ min^{-1}$). B: translational bursting is found when mRNA abundance is decreased tenfold ($s_A = 5\ min^{-1}$ and $s_R = 0.5\ min^{-1}$). C: given the conditions in A, now the transitions between active and repressed states are fixed to around one per hour, instead of 10 per minute ($k_{on} = k_{off} = 0.02$ per minute). This leads to transcriptional bursts in mRNA production, consequently affecting protein abundance.

explain the processes under analysis. The previous statement tries to make clear that the evolutionary essence of the systems here described has been always considered, although it lies deep in the background when specific details are treated.

2. Generation of periodic signals: calcium oscillations

The next chapter describes in detail a specific signal transduction pathway: the IP_3-mediated calcium pathway. After a brief introduction, a novel method to analyze this purely nonlinear system will be shown. This method ultimately allows us to get a detailed description of the complex molecular interactions present in this system, proving to be a powerful tool which may be useful for the analysis of other oscillating systems inside cells.

2.1. Introduction to calcium signaling

Together with phosphate ions, calcium ions play a main role as messengers in cellular signaling. Phosphorylation triggers the function of many proteins by binding and unbinding phosphate ions to specific domains. In a similar way, calcium interacts with specific proteins to change their shape and charge, and therefore to regulate their function inside the cell. The importance of calcium signaling is associated to the crucial role it plays in many cellular processes like oocyte fertilization and development, muscle contraction, cell proliferation, fluid secretion and metabolism [82, 83].

Calcium is just a chemical element and therefore it is unable to show conformational or chemical changes upon interaction with different stimuli. Nevertheless, the mechanisms which have evolved to use it as an information carrier in many different cellular processes, compensate that by generating complex spatio-temporal patterns, allowing cells to condensate a broad set of different external signals into a single carrier. At the same time, this lack of chemical flexibility implies that, unlike phosphate ions, calcium ions need to be somehow amplified to be able to interact with the much bigger proteins. This amplification is performed by calmodulin, a small protein which changes its conformation upon calcium binding to subsequently control other protein function and dimerization capabilities [84, 85]. Not only calcium pathways are governed by calmodulin, this protein can activate certain phosphorilation pathways [86], giving rise to a cross-link between both signaling mechanisms.

How do cells manage such an small molecule in a precise and reliable manner? To generate the broad spectra of different responses needed to encode a big amount of information in such a simple molecule, calcium is managed by pumping it in and out the cell and into the

different organelles which serve as calcium reservoirs. When a specific signal arrives to the cell, different reactions give rise to calcium release into the cytosol from the endoplasmic reticulum (ER) or the equivalent sarcoplasmic reticulum (SR) in muscle cells, activating the subsequent signaling cascade. There are two main processes which give rise to this calcium release: the voltage-induced calcium release and the IP_3-mediated calcium release [86].

2.1.1. Voltage-induced calcium release

Muscles work thanks to a propagating action potential which cordinates the responseof their cells. This mechanism is triggered by a voltage change in the cellular membrane Fig.(2.1): membrane depolarization opens the calcium-selective channels which allow the fast passage of calcium from the extracellular media to the cytosol. Increase in the cytosolic calcium concentration induces the calcium release from the ER (or SR in muscle cells) into the cytosol by opening the calcium-sensitive ryanodine receptors (RyR). This increase in calcium concentration ultimately leads to muscle contraction by interacting with troponin C and by binding to calmodulin to initiate the production of ATP through the glycogen metabolism. This fast voltage-induced calcium release allows skeletal muscle cells to respond rapidly or keep the right pace in cardiac tissue.

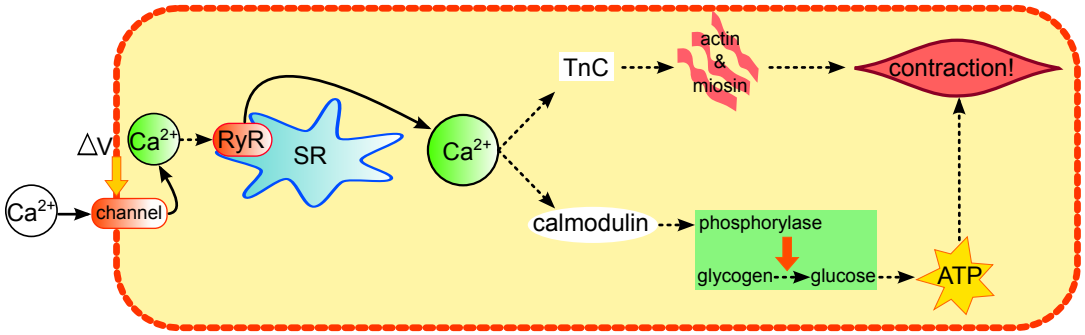


Figure 2.1.: Cartoon of the voltage-induced calcium release pathway, with solid lines representing calcium fluxes and dashed lines interactions between molecules: upon membrane depolarization ΔV, specific calcium channels open, allowing the passage of calcium into the cytosol. This calcium ions trigger more release from the SR via the ryanodine receptors (RyR). The new boost in calcium concentration activates two parallel processes: on the one hand it interacts with troponin C (TnC) to start muscle contraction while on the other hand it binds to calmodulin which activates the glycogen metabolism to generate ATP.

2.1.2. IP3 as a second messenger

Upon arrival of different stimuli to specific receptors lying in the cell membrane, the IP_3-mediated signaling pathway is activated. Many different signals activate this pathway [24], and therefore during this process cells multiplex them into a single carrier: cytosolic calcium

ions. To be able to perform this complex function, calcium concentration in the cytosol exhibits a variety of different behaviors, with different long-term and short-term spatial and temporal patterns [87]. Perhaps the most eye-catching behavior of this mechanism is the production of sustained calcium oscillations: cells seem to be translating amplitude-modulated (AM) signals into frequency-modulated (FM) ones, showing evidence that complex pathways are working inside them. The shape of these oscillations resembles more a train of spikes rather than a sine-like wave, and therefore we will sometimes say that the system is 'spiking' instead of oscillating.

At first glance we notice two qualitative properties of this system:

- Depending on the external signal, the spiking pattern changes, with a wide spectra of different inter-spike periods available [24]. An example of this is provided in Fig.(2.2), where Hela cells are stimulated with histamine and ATP.

- Given a specific signal, the inter-spike period also depends on ligand concentration. For example, in Hela cells we observe that, when stimulated with an increasing histamine concentration, the spiking frequency increases Fig.(2.3)). Therefore, not only the type of stimulus is multiplexed, but also its concentration is being implemented in the response.

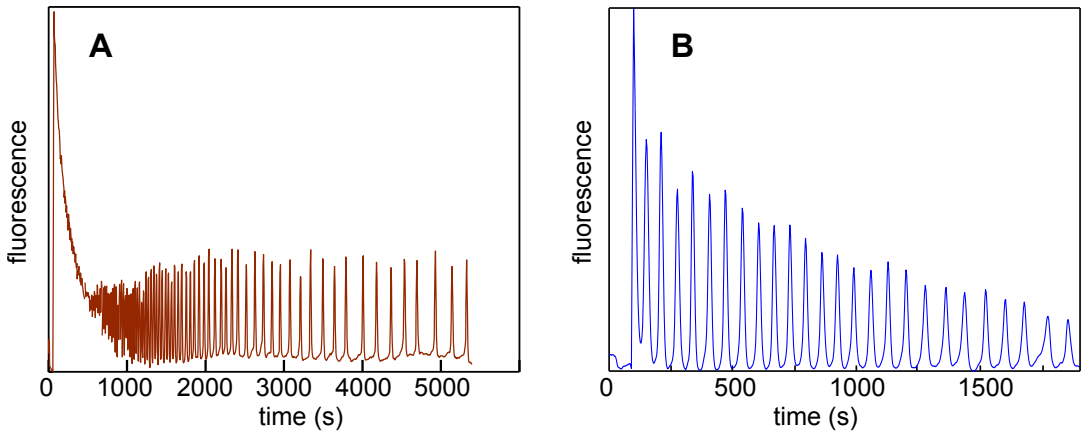


Figure 2.2.: Response to ATP and histamine: trajectories showing stereotypical calcium behavior in Hela cells when they are stimulated with 10 μM ATP (A) and histamine (B). Notice how the shape of both trajectories differs and how calcium spiking is faster for ATP stimulation.

It is clear that the set of reactions that are induced upon activation of this pathway have evolved in a strongly non-linear way which allows it to exhibit sustained oscillations with relatively constant frequency. How is exactly the intracellular machinery giving rise to this complex behavior? While the voltage-induced calcium release mechanism is relatively well known, the details of the mechanism governing calcium signaling mediated by IP_3 are still unclear, with some authors suggesting the existence of certain feedbacks while others proposing different interactions. Nevertheless, although the specific interactions between them are still to be confirmed, there is a relatively good knowledge about the components intervening

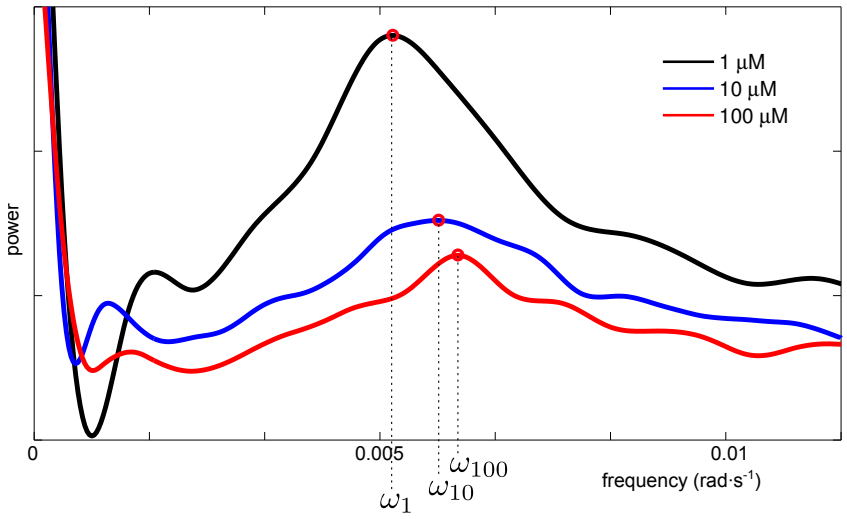


Figure 2.3.: Spiking frequencies for different histamine concentrations: mean power spectra obtained by averaging individual spectra calculated by Fourier transforming experimental time series obtained with Hela cells under 1, 10 and 100 μM stimulation. The spiking frequency increases with ligand (histamine) concentration.

in this process. They can be divided in three main types:

1- Receptors, pumps and channels.

2- Proteins and other second messenger molecules.

3- Calcium pools.

Among the first class components we find membrane receptors like G protein-coupled membrane receptors (GPCR), pumps moving calcium from the inside towards the outside of the cell, pumps in the ER surface which pump calcium inside it like sarcoendoplasmic reticular calcium ATPases (SERCA pumps), and receptors in the ER surface which sense second messengers and allow calcium release into the cytosol. In the second group we have inositol (1,4,5) trisphosphate (IP_3), the main second messenger molecule involved in this pathway, together with other molecules involved in IP_3 synthesis and degradation. Finally, we have two main calcium pools: the cytosol itself and the ER (with some authors proposing the mitochondria as an important pool too)

A schematic of this pathway is provided in Fig. (2.4). After a signal binds to a receptor in the cell membrane, the basic course of events would be as follow: upon signal binding to a GPCR the G-protein diffuses through the membrane and activates phospholipase C (PLC) which subsequently hydrolizes phosphotidylinositol-(4,5)-biphosphate (PIP_2) into dialcylglycerol and IP_3. IP_3 diffuses through the cytosol and binds to IP_3 receptors (IP_3r) found in the ER surface. This binding of IP_3 to an IP_3r opens a channel and calcium stored in the ER is released into the cytosol. Calcium is constantly pumped back into the cytosol via SERCA pumps, and outside of the cell by channels in cell membrane, while some leak of calcium from the extracellular media into the cytosol may be present too. IP_3 is degraded by

a phosphatase and subsequently converted to PIP_2.

As we have already stated, experiments show calcium spikes upon stimulation with certain signals, although given the basic picture described above, no oscillations would be expected: the external stimulus activates the synthesis of IP_3, which gets to a certain equilibrium concentration. IP_3 activates a process in which calcium moves from the ER to the cytosol until the rate of calcium going out equals the rate of calcium going in and an equilibrium concentration is reached. To get this system oscillating more ingredients are needed: a sort of negative feedback on calcium release which would rapidly reduce cytosolic calcium concentration once a threshold is reached. Different authors have proposed a broad set of additional interactions and mechanism to get the system oscillating:

Meyer and Stryer [88] proposed calcium sequestration by mitochondria (although it could be a leak outside of the cell) to be responsible for the oscillations. Goldbeter et al. [89], Atri et al. [90] and Li et al. [91] developed different models which achieved calcium oscillations by including a negative feedback from the cytosolic calcium to the IP_3r, which blocked calcium release from the ER once a certain calcium concentration is reached. Then, SERCA pumps move the cytosolic calcium back into the ER again, reducing the concentration until IP_3r's are activated again. Finally, Sneyd et al. [92] proposed a completely different model in which the dynamics of the IP_3r are described in detail and are enough to give rise to oscillations.

In these studies, all these mechanisms are translated into mathematical models which allow for a good characterization of the hypothesis. When a model is developed assumptions are made not only for the interactions between system components (pumps, receptors, messengers...), but also for the dynamics of those components alone (susceptibilities, degradation mechanisms..). This gives rise to models which exhibit different behaviors even though their interactions are the same.

Some studies have focused on the validation of the interactions between components, the topology of the system: among the models mentioned above, some of them include a feedback from cytosolic calcium to PLC, which is ultimately translated in an effective feedback from calcium to IP_3 [88]. This specific feedback has been studied in different works: James Sneyd and coworkers [68] monitored the behavior of the calcium oscillations when IP_3 was suddenly added to the cells. At the same time, they took 13 different models from previous authors and modified them to get a class 1 (with no feedback) and class 2 (with feedback) version of each one. The equivalent experiment of sudden IP_3 addition was performed in-silico over the different mathematical models and the results were compared with the experiments. The study was performed over two different cell types, pancreatic acinar cells (PAC) and airway smooth muscle cells (ASM), and concluded that the long period oscillations found in PAC depended on IP_3 oscillations, whereas the shorter-period oscillations in ASM do not. Another work by Toru Matsu-ura et al. [93] monitored IP_3 *in vivo* and concluded that a feedback on IP_3 may not be present in Hela cells.

Apart from the topology, many other details are still unknown in this system: how do

IP_3r's work? how is the dynamics of the SERCA pumps? is calcium sequestration by mitochondria a must to obtain oscillations? This questions could be answered if a more precise method to compare models and experiments is developed, allowing to select the model which better fits the experimental results and therefore whose dynamics gives a better description of the real system. In the next sections we describe our hypothesis of such a method, with some promising results emerging from it.

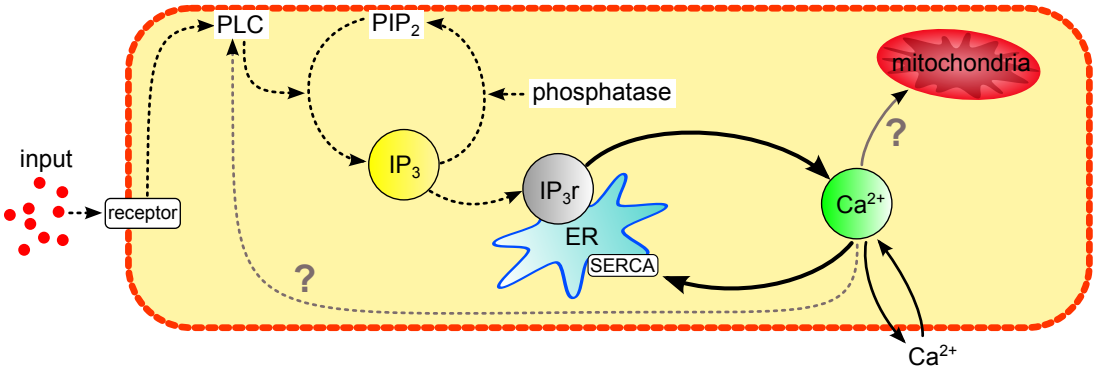


Figure 2.4.: Cartoon of the IP_3-mediated pathway, with interactions indicated by dashed lines and calcium fluxes by solid ones. Upon the arrival of an external signal to a G protein-coupled membrane receptor (dark green), Phosphotidylinositol-(4,5)-biphosphate (PIP_2) is hydrolized by its phospholipase C (PLC) to dialcylglycerol (not shown) and IP_3. IP_3 molecules diffuse through the cytosol to bind to specific receptors (IP_3r) in the ER membrane. There, a calcium channel is opened to allow the release of calcium from the ER into the cytosol. IP_3 is degradated by IP_3-phosphatase to give rise again to PIP_2 and calcium is constantly pumped into the ER by the SERCA pumps. Two main unknown interactions and fluxes are indicated by gray lines: calcium sequestration by mitochondria and calcium feedback on IP_3 production.

2.2. Cellular interrogation

So far we have provided a brief introduction to calcium signaling, showing how many vital cellular processes are regulated by its dynamics. Two main pathways have been described: the voltage-mediated and the IP_3-mediated pathway. The research presented on this chapter is focused on the latter, and therefore, we have described it with more detail, highlighting the current uncertainties on the different interactions and the specific mechanisms underlying the observed behavior.

We can provide two main reasons to study the IP_3-mediated pathway: first, although it is a cornerstone in many cellular processes, helping to process different signals and cellular responses, it is still not well understood. Second, it represents a paradigmatic example of an oscillating cellular system, providing a good field of action where new techniques to study

such complex systems can be developed and tested.

To address this problem we have developed a 'cellular interrogation' protocol which allows us to better understand the underlying mechanisms with a minimal intervention inside the intracellular machinery. The idea is to 'ask' the same questions to real cells and to different mathematical models which propose different mechanisms to describe the system under study. Then, the 'answers' provided by the models can be compared with the experimental one, allowing to accept or reject the descriptions provided by the different models. A first naive approach to this process is described in Fig. (2.5).

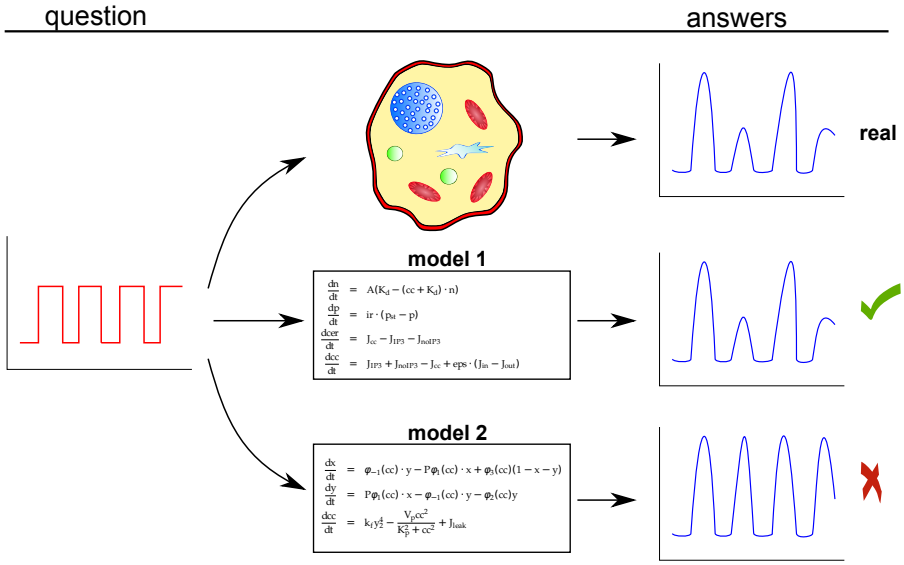


Figure 2.5.: Naive cellular interrogation: the same question, in the form of a complex input signal, is asked to real cells and different models. The answers are compared, allowing to discard some models and keep others.

We have called the process described in Fig. (2.5) 'naive' because it would only work in an ideal world. In the real world things get more complicated: first, during an experiment not all the cells behave in the same way. Cell to cell variation gives a broad set of different responses to the same signal, even in clonal cells cultured under the same conditions. Second, the responses given by the models strongly depend on the specific parameters under which they are tested. Although the parameter values given by the original authors of the models are always biologically inspired, there is always a certain degree of uncertainty over their real values. Moreover, since these parameters are sometimes effective parameters which account for more complex mechanisms, their value may change during the cell cycle. Therefore, at this point a direct comparison between the behavior of real cells and theoretical models under specific parameters seems useless.

To get rid of cell to cell variation and parameter uncertainty we have to take a longer route. The novel approach we have developed is based on finding statistical 'fingerprints' or 'signatures' which hopefully would be uniquely associated with each model, regardless the

parameter values. In a similar way, an experimental signature could be computed by doing the same statistical analysis over many experiments. To finish, signatures can be compared and models can be accepted or rejected depending on how similar their signature is to the experimental one.

The data needed to go through this process is obtained as follows:

- During an experiment, the natural cell to cell variation gives the necessary data to generate a signature: for many experiments calcium at the single-cell level is monitored to obtain a broad set of different responses. Then, a statistical analysis of those responses is performed and a 'experimental signature' is generated.

- To simulate the observed cell to cell variation during the model analysis, the response of each model is computed for different sets of random parameters. With this random-parameter approach we expect to mimic the fact that cells may be in different states of their cycle , have different shapes, different number of membrane and IP_3 receptors and so on. Then, the very same statistical analysis performed over the experimental data is applied to the results obtained with each of the models, therefore generating individual signatures for each of them Fig (2.6).

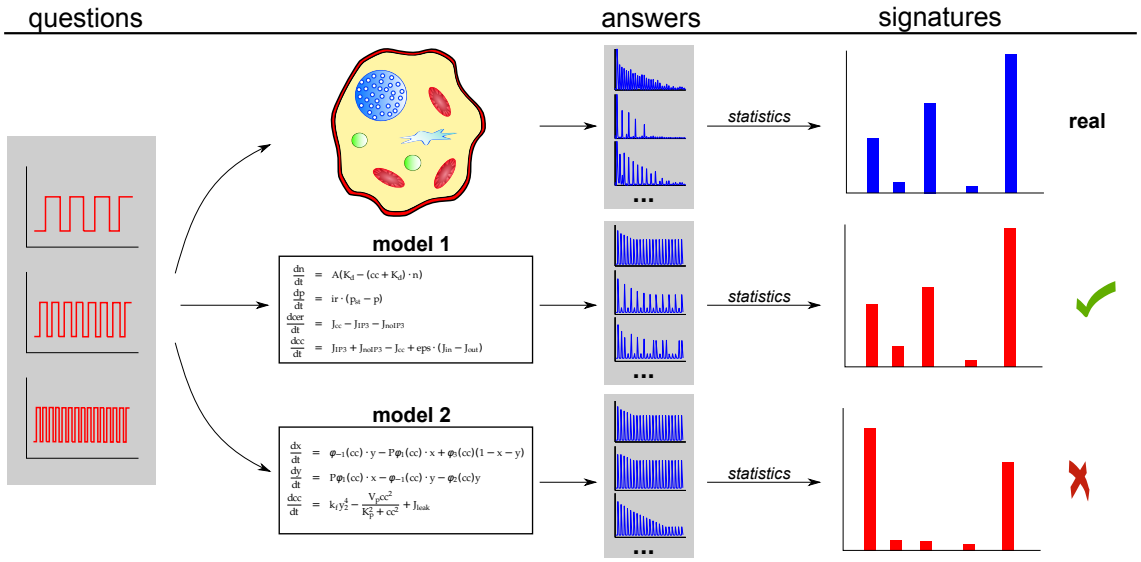


Figure 2.6.: Real cellular interrogation: many questions, in the form of complex input signals, are asked to real cells and different models with different parameter sets. Multiple answers are obtained and analyzed. A statistical 'signature' is generated for each model and the experimental data. Signatures are compared and certain models are rejected while others accepted based on the similarity of their signature with the experimental (real) one.

Novel experimental and computational approaches have been developed to quantitatively characterize the behavior of both models and cells:

- Microfluidic devices give us the possibility to stimulate cells with well controlled signals.

We have designed and developed novel microfluidic devices which allow us to stimulate the oscillatory machinery of calcium with complex input signals while the calcium response is monitored via fluorescence. This complex signals are the 'questions' asked to the cells, while the calcium response would be the 'answer'.

- Previously published mathematical models proposing different mechanisms for our system provide us with a pool of different hypothesis to be tested and compared with the experimental results. A computationally intensive approach has been taken to simulate the microfluidic stimulation process within each different model for a broad set of different parameters. Therefore, the same 'question' asked to real cells would be asked to each model.

- Finally, a *pattern recognition* routine has been developed to analyze the trajectories obtained for both models and experiments. The specifics of the pattern recognition define the 'language' in which 'signatures' would be translated to.

All the experiments have been performed on Hela cells and therefore, given the heterogeneous responses already described for different cellular types [68, 82, 83], our results about the calcium machinery should not be directly generalized to other cellular types. Nevertheless, exactly the same method could be used to characterize calcium dynamics within many different cell types.

Although it is still to be tested in other systems different than calcium dynamics, this method may provide a general recipe to unveil the molecular details of many different molecular networks which exhibit natural oscillations and whose study has been always difficult due to their intrinsic non-linearity.

All the work described in this chapter has been done together with Natalie Andrew, Frederick Chang, Florian Gnad, Daniel Gibson and Jeremy Gunawardena, from the Gunawardena Lab at Harvard Medical School.

In the next sections the experimental and computational methods will be described in detail, followed by a a description of the results and a final discussion.

2.3. Methods

In the next section we will give an overview the tools used during the 'cellular interrogation' process. First, an introduction on the experimental methods is provided, describing the fabrication of microfluidic devices, the data acquisition protocol and the segmentation of the obtained images. Then, computational methods are described: cell to cell variation implementation, simulation of microfluidic experiments and statistical analysis.

2.3.1. Experimental

More than 50 years ago Leo Szilard invented the chemostat [94, 95]. This device allowed to control the growth of microorganisms by controlling the input rate of a limiting nutrient and the dilution rate of the culture. Following Szilard's idea, microfluidic devices were developed when lithography techniques started to be applied to create patterns *in silico*n based polymers [96]. This technique allows to build devices in which fluids can be confined in the scale of microns, giving researchers in cell biology the possibility to precisely control cellular media at the actual scale of cells. The power of this tool is only limited by the ability of scientists to design new devices and to propose new applications [97–99].

For our research, we have designed and built novel microfluidic devices to be able to generate time-dependent signals to stimulate adherent cells. The design creates an almost laminar flow over the surface through which cells are distributed, and therefore it guarantees that all of them are being stimulated with the desired signal. During stimulation, the response is monitored, obtaining a real-time stimulus-response relation which can be subsequently analyzed.

Microfluidic device fabrication and operation

Microfluidic experiments have been performed by Natalie Andrew, Frederick Chang and Daniel Gibson. The whole process involves device design and fabrication, cell culture and data acquisition.

Microfluidic devices consist of two layers: the flow layer and the control layer. The flow layer contains the channels through which the cellular media will be flowing during the experiments, while the valves controlling the flow are found in the control layer. Both layers are made of PDMS, which is shaped using the silicon masks shown in Fig(2.7 A-B) printed at 20,000dpi (CAD/Art Services, Inc.) yielding 10 micron minimum feature resolution. Mixed and degassed PDMS is placed on the masks and spun to get the desired thickness. The resulting assemblies are baked at 65ºC for at least one hour. Then, the control layer is peeled off and cut in the six individual pieces which would yield to the six different devices. Each piece is bound to the flow layer by plasma etching both pieces. This assembly is then baked again at 65ºC for another hour. After baking, the flow layer is carefully cut around the

perimeter of the six control layer pieces, and peeled from the flow-layer master. The last step to get the device ready is to punch in it holes where the control and flow tubes would be attached.

How do valves work? When the flow and control layers are bound, the little chambers formed in the control layer -red in Fig(2.7 C)- intersect the channels of the flow layer -blue in Fig(2.7 C)-. When pressure is increased inside one of those little chambers, its bottom bends, blocking the channel bellow and therefore stopping the flow. Now, by closing and opening the different valves in a coordinated fashion we can choose the source from which the flow arriving to the cell chamber comes from Fig(2.7C).

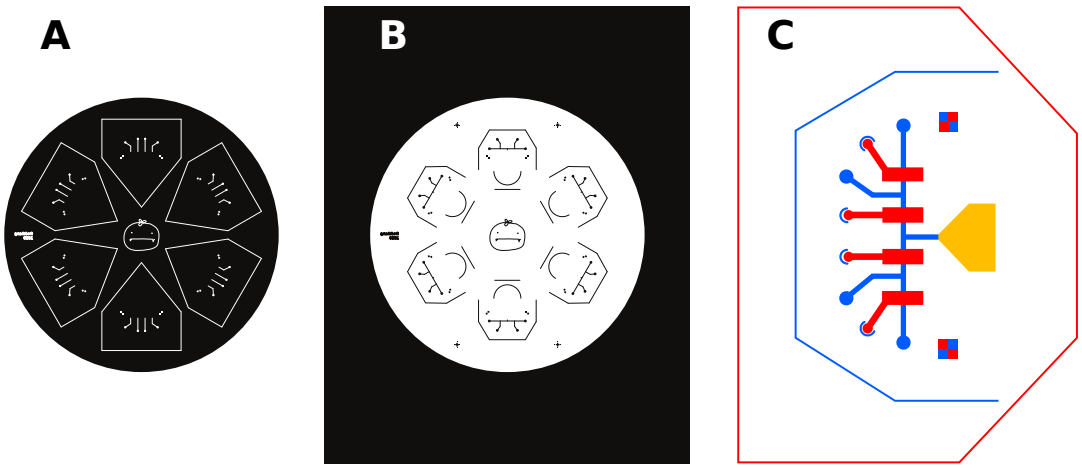


Figure 2.7.: Microfluidic device design: control (A) and flow (B) masks are used to fabricate six devices at a time. When both layers are bonded together (C), the control layer (red) overlaps with the flow layer (blue). Marks are used to align both layers. The cell chamber (yellow) is found at the end of the flow passages. Control and flow tubes are attached to the holes made at the circular structures.

Although an initial design in which cells where grown inside a closed chamber was tested, we obtain better results when the experiments are performed using a chip-in-a-dish configuration. The device is glued to the bottom of the dish and the original chamber is chopped off. Then, cells are cultured at 37ºC with 5% CO_2 for 24 hours inside that dish before the experiment.

Experiments are performed by connecting different tubes to the device. Some flow tubes contain buffer while others transport the desired stimulus. In our case, different histamine concentrations have been used as input. Water filled tubes are connected to the computer-controlled pneumatic actuators which would open and close the valves in the device. A *java* program allows us to control the state of the valves, and therefore, to generate the desired square waves of histamine concentration, with a precisely controlled period and pulse width.

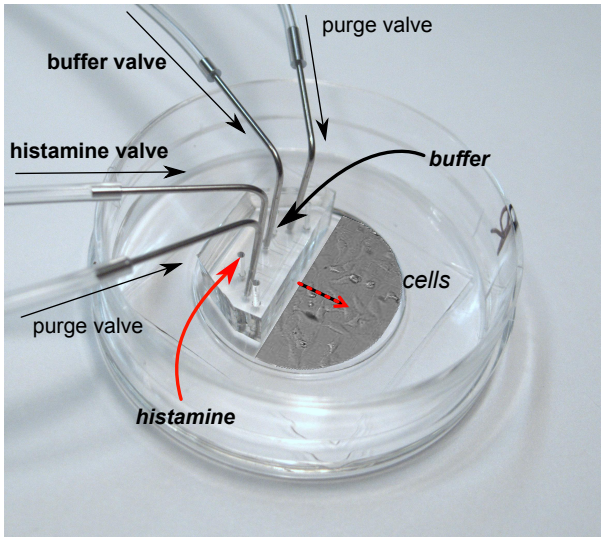


Figure 2.8.: Picture of a microfluidic device: histamine and buffer solutions are pumped in at 5 psig with tubes attached to the indicated holes (tubes not shown). Valves are controlled with compressed air at 25 psig: with the purge valves closed, an alternate opening and closing of the buffer and histamine valves stimulates the cells with controlled histamine pulses.

Determining the system time-scale: constant stimulation

When stimulated with a constant concentration histamine step, most of the cells exhibit decaying amplitude and increasing period cytosolic calcium spikes Fig (2.9). We have performed experiments with $1\mu M$, $10\mu M$ and $100\mu M$ constant histamine concentrations Table(2.1), observing that the general tendency is that the higher the histamine concentration, the smaller the typical inter-spike period Fig.(2.3).

experiment	concentration (μM)	# segmented cells
1	100	262
2	10	286
3	100	229
4	10	249
5	1	278
6	10	535

Table 2.1.: List of constant stimulation calcium experiments: in each experiment, histamine concentration kept constant.

$10\mu M$ stimulation seems to give a good cell response, being the chosen concentration for the main 'cellular interrogation' experiment, where cells are stimulated with histamine pulses of various frequencies and pulse widths.

To have an estimation of the time-scale cells are working in, a first analysis is performed

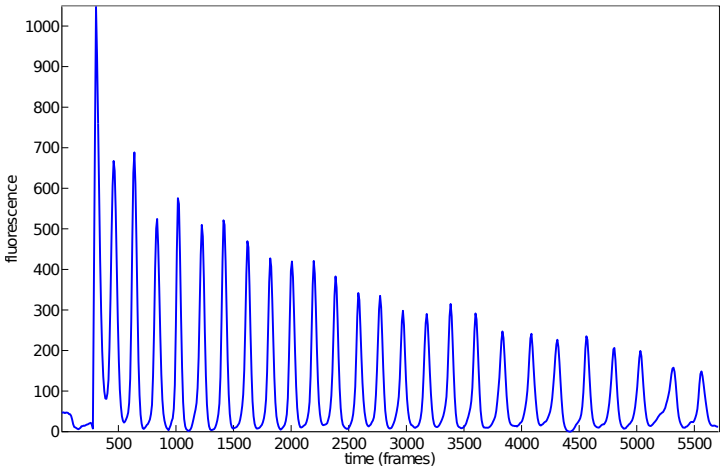


Figure 2.9.: Response to constant histamine stimulation ($10\mu M$). After an initial high spike, spike heights decrease monotonically. At the same time, the inter-spike period increases.

for $10\mu M$ constant stimulation. Although cell to cell variation is present and the inter-spike period is not constant within each cell, we can compute an average spiking period $T_{av.}$ for this concentration. Averaging over 3 different experiments, for a total of 192 responding cells, we compute an average spiking period of $T_{av.} = 130 \pm 40s$ Fig (2.10). This average period helps us to fix the different frequencies and pulse widths which would be used when a more complex square-wave stimulation is applied.

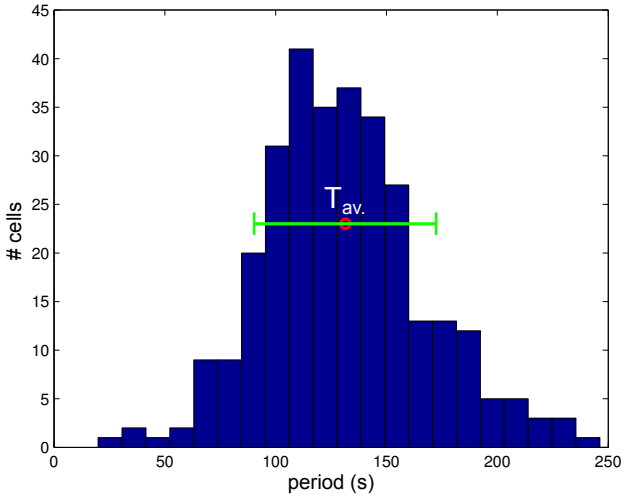


Figure 2.10.: Experimental inter-spike period distribution calculated from 3 independent experiments and a total of 192 responding cells. The average interspike period is $T_{av} = 130 \pm 40s$

Interrogating cells: pulsed stimulation

Once microfluidic devices have been tested and a nice working experimental protocol has been established, the main experimental part of the 'cellular interrogation' procedure consists on stimulating cells with complex signals which would work as 'questions' asked to them. A pulsed stimulation signal seems to be a good candidate to play this role: fixed the amplitude ($10\mu M$), it can be tuned in both frequency and pulse width, allowing for a flexible yet simple signal generation. In addition, this complex stimulus is still simple enough to be easily modelized when the different mathematical models are subjected to the same conditions. A description of the signal characteristics is provided in Fig. (2.11).

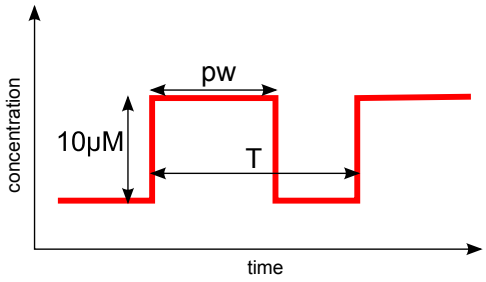


Figure 2.11.: Histamine stimulation: $10\mu M$ square wave of period T and pulse width pw.

We have performed 15 independent experiments for different input periods (T) and pulse widths (pw), as shown in the Table (2.2).

In the experiments, histamine pulse widths pw range from 38 to 5 seconds. Comparing these numbers with the previously calculated average period for constant stimulation $T_{av.} = 130 \pm 40s$ helps us to locate in which dynamic region of the system are the *in vivo* experiments done, setting the framework in which *in silico* experiments should be performed for a good comparison. For each pulsed stimulation experiment, the ratio L between $T_{av.}$ and pw is used as a reference value for this purpose. L and its associated experimental error ΔL are calculated using eq. (2.1), giving values of L between 0.29 and 0.04 eq. (2.2)

$$\begin{aligned} L &= \frac{pw}{T_{av.}} \\ \Delta L &\simeq L \cdot \frac{\Delta T_{av.}}{T_{av.}} \end{aligned} \tag{2.1}$$

$$\begin{aligned} pw = 38s &\rightarrow L = 0.29 \pm 0.09 \\ pw = 5s &\rightarrow L = 0.04 \pm 0.01 \end{aligned} \tag{2.2}$$

The typical cell response to a pulsed stimulus is plotted in Fig. (2.12). Calcium spiking always occurs when a histamine pulse arrives. When signals are too fast to be followed by

experiment	T (s)	pw (s)	# segmented events	# accepted cells
1	40	15	108	98
2	40	15	153	137
3	60	22	575	449
4	20	5	460	417
5	60	22	568	461
6	60	38	299	229
7	20	5	649	595
8	40	10	461	247
9	40	10	402	334
10	40	25	367	366
11	60	38	364	219
12	60	38	379	276
13	25	15	183	176
14	25	15	411	386
15	25	15	173	123

Table 2.2.: List of pulsed stimulation calcium experiments: in each experiment, histamine oscillates between 0 and $10\mu M$ concentration, with period T and pulse width pw.

the system, it starts skipping input pulses. Not all the cells behave in the same way, with some of them being able to follow some fast signals that produce a lot of pulse skipping in others.

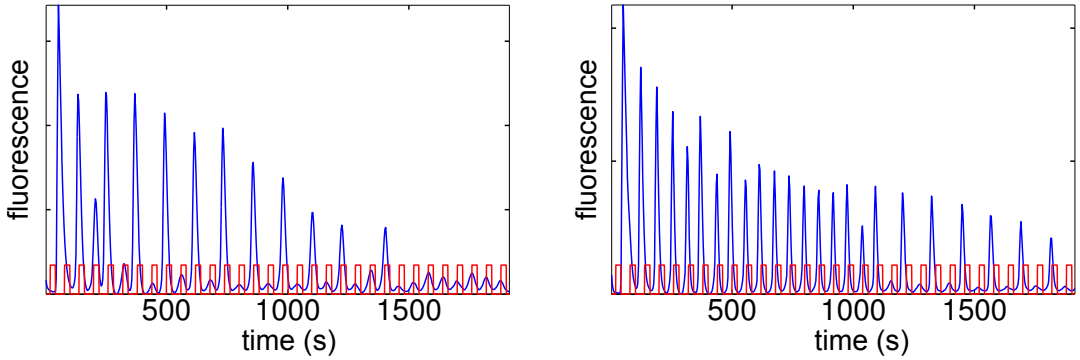


Figure 2.12.: Two typical skipping patterns in an experiment: in the same experiment, two cells show different responses for a pulsed stimulation. Calcium spikes (blue) occur always when a histamine pulse arrives (red). For fast enough signals, some cells (left) skip some pulses, showing little calcium response. On the other hand, other cells (right) are able to better follow the pulsing stimulus. Experiment performed for $10\mu M$ histamine pulses of 60 s period and 22 s pulse widths.

Image analysis & event acceptance-rejection

Once the experiments have been performed, calcium is computed at the single cell level. Each image sequence is segmented, allowing us to measure the fluorescence intensity at each segmented event. A macro for *ImageJ* has been written to perform segmentation in a systematic manner Fig. (2.13). Integrated density -the product of the event area and its mean grey value- is used as a measure of the response.

Once the experimental data have been processed with *ImageJ*, we use *Matlab* to analyze the resulting trajectories. Before going on with the statistical analysis which would lead to the final signature of the experiments, the trajectory of every segmented event is analyzed in order to discard the ones which may correspond to little bubbles or dead cells in the field of view rather than actual responding cells. The criteria we use on this event acceptance-rejection process is just to keep any single trajectory whose response (spikes) is entrained with the input stimulus. To do so, we follow the next three steps:

1.- For each trajectory a background correction is applied [100] and peaks are found [101].

2.- Peak positions are used to compute an average period $T_{av.}$.

3.- If $T_{av.} \pm 0.3T_{av.}$ is above or bellow the fixed input period, the event is discarded.

Table (2.2) gives the number of segmented events as well as the final accepted ones for each of the 15 experiments. These results are plotted in Fig. (2.14). One should notice that the number of accepted cells once the undesired events are discarded is always high (well above 50%), meaning that the overall experimental protocol is well refined.

To fine tune the background correction and peak finding parameters before running the analysis described above, we have written a visual *Matlab* GUIDE program which loads each experiment data and helps us to go through and visualize each event trajectory, showing in real-time how the background is being corrected and the peaks selected when their thresholds are varied.

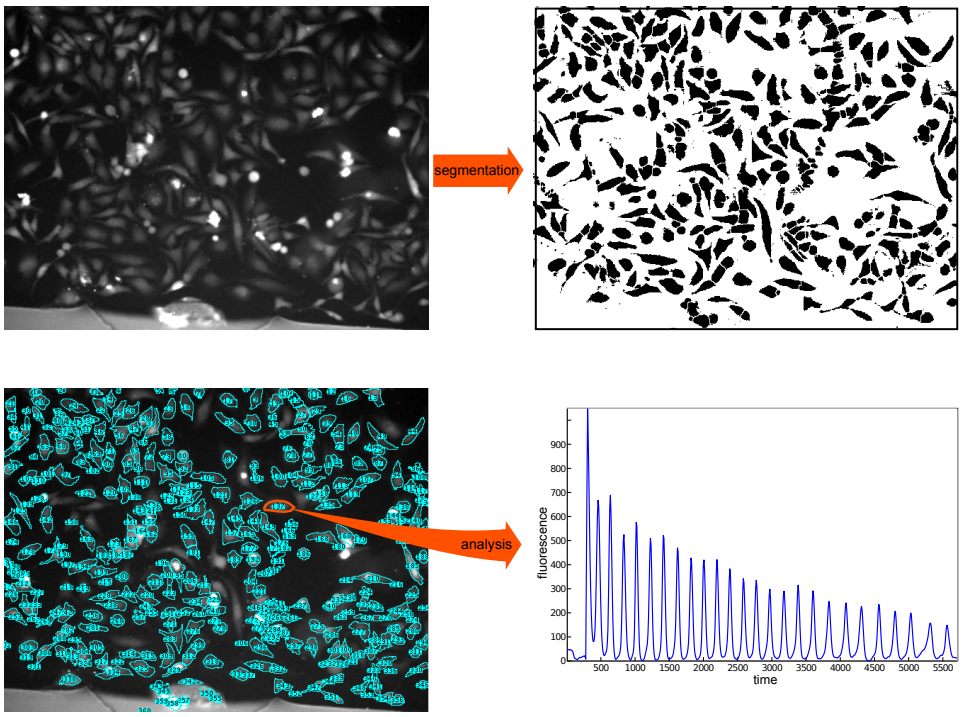


Figure 2.13.: Segmentation process: for each experiment, an average image of all the frames in the movie is created. The averaged image is converted to a mask which is segmented, creating a multiple region of interest (ROI) where cells are selected. For each ROI element, fluorescence integral density is measured along the whole movie.

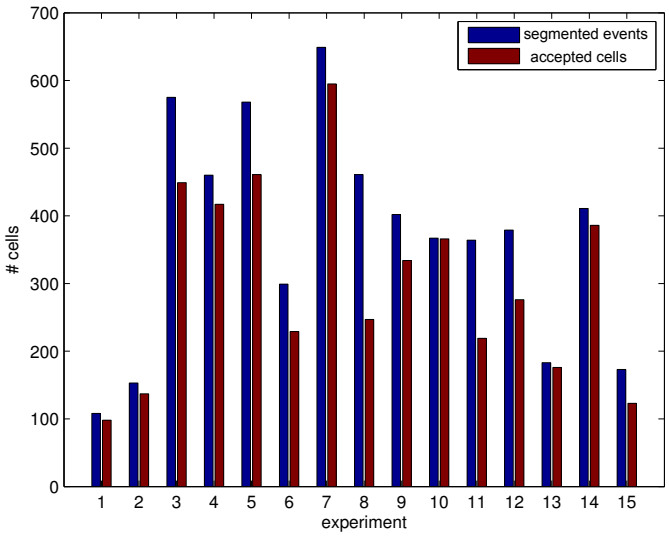


Figure 2.14.: Accepted-rejected events: for each experiment, the number of segmented events (blue) is plotted together with the number of accepted ones (red).

2.3.2. Computational

The 'cellular interrogation' procedure consists of two main steps: the experimental analysis of real cells and the simulations performed over different mathematical models describing the calcium oscillations system. In this section we provide a description of the analytical and computational methods used in the second step of the process.

The final step of the process will be to compare the behavior exhibited by different models extracted from the literature with the one shown by real cells, in order to keep the models which better describe reproduce the experimental results. Therefore, the first step here would be to choose the initial set of mathematical models properly.

Models

In principle it could happen that none of the models under study fits the experimental behavior, giving no final answer. To avoid that, we have chosen eight previously published models describing relatively different mechanisms Table (2.3). We expect that having a wide spectra of hypothesis would increase our possibilities of finding a 'correct' model for our system. Therefore, one reason for choosing the IP_3-mediated calcium pathway is the abundance of published models describing its dynamics. In the introductory section of this chapter we have described some of the interactions which are still not clear in this system. One of those interactions is the feedback from cytosolic calcium to IP_3. James Sneyd and coworkers [68] classify most of the already proposed models into two classes: class 1 models, where no IP_3 feedback is present, and class 2, where it is [68]. Following their criteria we have chosen models of both classes, extracting them directly form Sneyd's paper. Among these models we have the ones proposed by Atri et al. [90] and by Li & Rinzel [91], in the modified versions presented by Sneyd. In addition, two traditional papers on calcium dynamics are studied too: Meyer & Stryier [88] and Goldbeter et al. [89]. Finally, a slightly different model proposed by Sneyd himself is analyzed in its 2 and 3 states versions. A list of all these models, together with the short name we have used all over this chapter for each of them and the corresponding references is provided in Table (2.3).

Once models have been chosen, the next important step of the present work is to subject them exactly the same experiment we are conducting over living cells. To do so, we have to somehow account for cell to cell variation, as well as to be able to locate the *in silico* experiments on the same dynamical region of the *in vivo* experiments.

Cell to cell variation is simulated by exploring the behavior of each model for a broad ensemble of different randomly generated parameters sets and initial conditions. Pulsed stimulation simulations are performed in a dynamical range similar to the one calculated for the experiments in eq. (2.2).

Model	Short name	references
Meyer & Stryer	Meyer	[88]
Goldbeter et al.	Goldbeter	[89]
Class 1 Atri et al.	Atri1	[68, 90]
Class 2 Atri et al.	Atri2	[68, 90]
Class 1 Li & Rinzel	Li-Rinzel1	[68, 91]
Class 2 Li & Rinzel	Li-Rinzel2	[68, 91]
2 state Sneyd et al.	SneydLeBeau2	[92]
3 state Sneyd et al.	SneydLeBeau3	[92]

Table 2.3.: List of analyzed calcium models.

Simulating cell to cell variation: random parameters

To account for cell to cell variation, sets of random parameters (PS) and initial conditions (IC) are generated. Each single parameter is randomly generated following a uniform distribution between 0 and 5 times its original value (the one given in the original publication), applying the same criteria to the IC. We assume that this range is enough to account for the observed cell to cell variation.

Since we are studying the molecular mechanisms which give rise to calcium oscillations, only PS's and IC's which place the models in their limit cycle regime are accepted. Fig. (2.16) shows that the range chosen to generate random parameters is reasonable, since some parameters prevent the system to oscillate when they get values above 5 times the original one.

A flux diagram describing the logic of the program which generates the random sampling is provided in Fig. (2.17). The program is written in *Fortran* to speed-up calculations and therefore improve statistics. The systems of ordinary differential equations (ODE) are solved with the common 4^{th} order Runge-Kutta method. Given a specific set of randomly generated initial conditions and parameters, the program looks for oscillations in the system: when an oscillating set is found (this is done simply by finding points in the trajectory which are higher than their immediate surroundings), the average period T_{nat} of the system under this conditions is computed and saved together with PS and IC. The final output of the program consists of a file containing 20,000 sets of random PS and IC, with their associated T_{nat}'s.

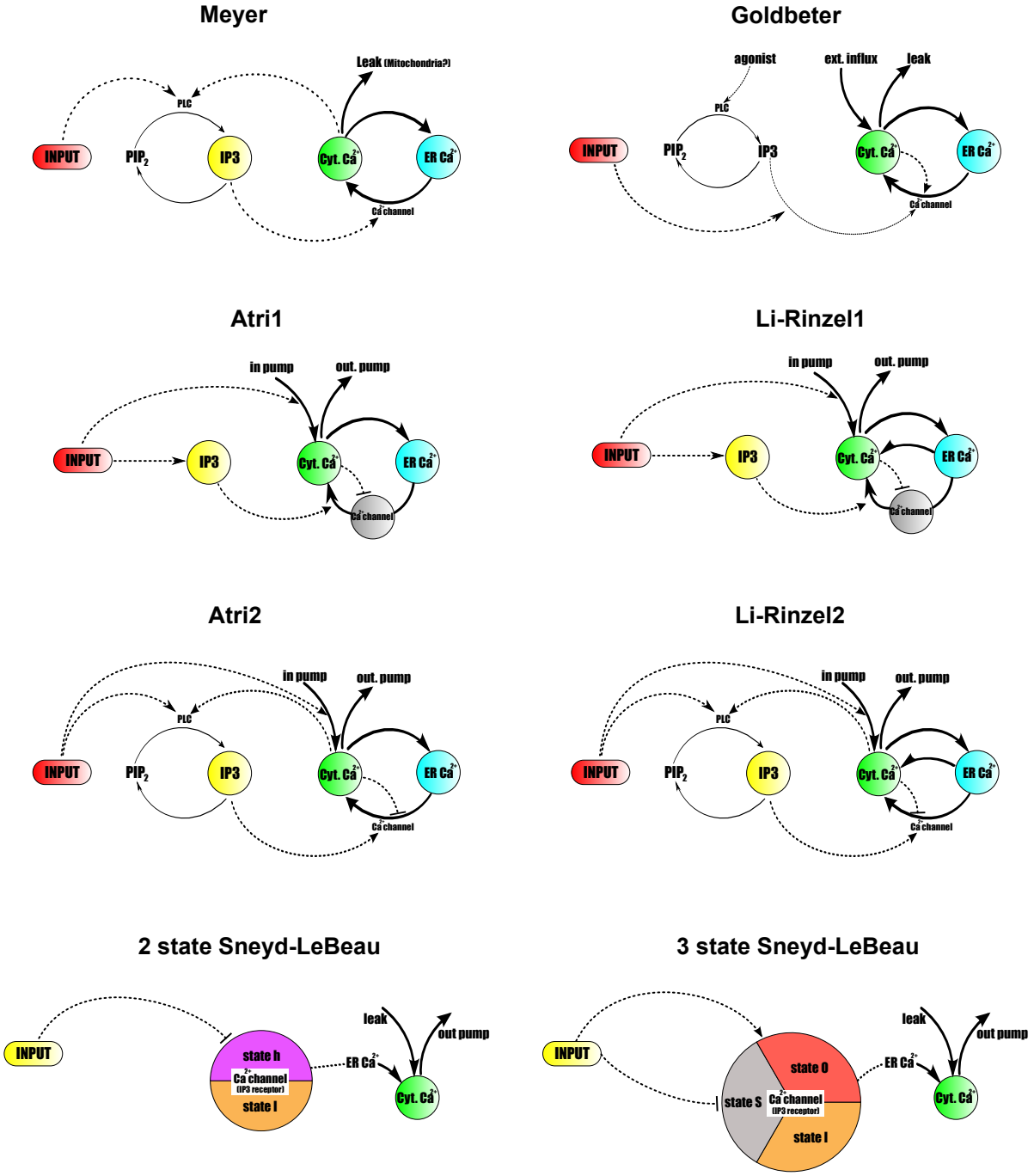


Figure 2.15.: Topology of the different models under study. Dashed lines represent activation-deactivation interactions, while full lines state for calcium fluxes. Circles indicate species whose dynamics are explicitly included in the model. Sneyd-LeBeau's models exhibits a slightly different scheme, with the big multi-colored circle representing the different states of the IP_3 receptor.

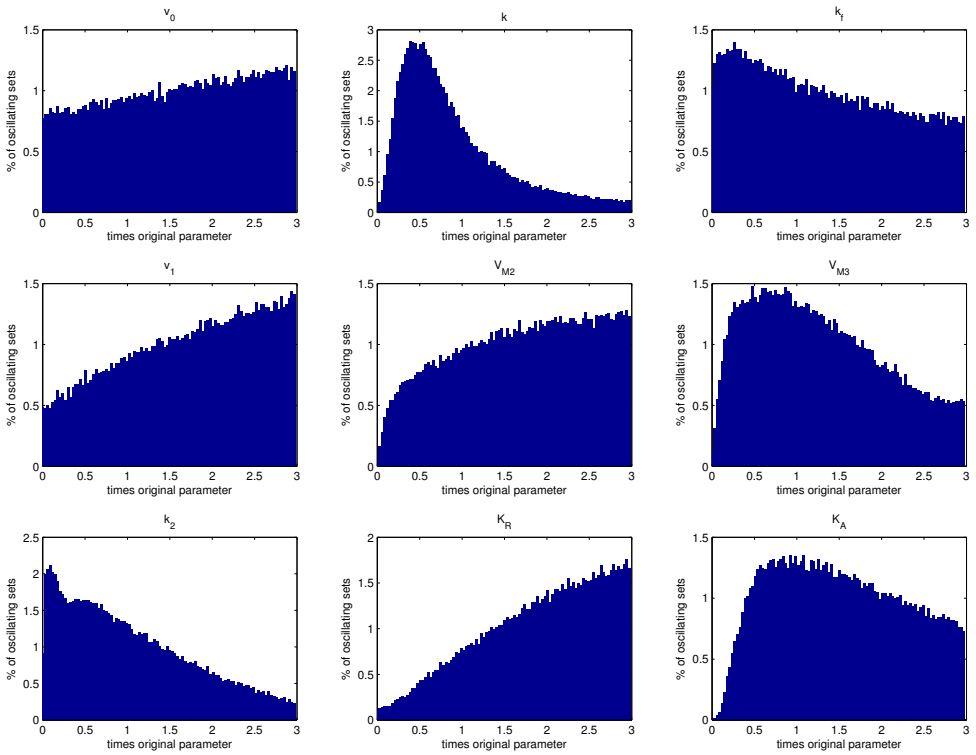


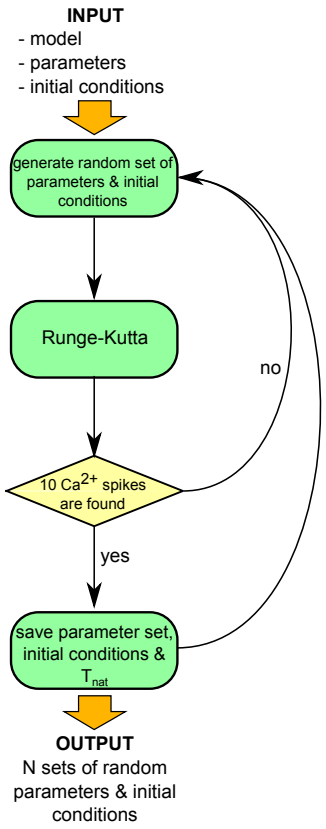
Figure 2.16.: Parameter distribution for the Goldbeter model [89]: histograms showing the distribution of each random parameter once parameter sets which prevent the system to oscillate have been discarded. 'y axis' indicates the percentage of sets exhibiting a given parameter value. Since multiple parameters are involved, the whole distribution for a single parameter does not sum 100. Notice that parameters k and k_2 limit the ability of the system to show oscillations when they go above 3 times their original value. Although the example provided corresponds to Goldbeter model, similar results are obtained for the rest of the models. This histograms have been generated with 20,000 sets of parameters.

Simulations with pulsed stimulation

Like in the experiments, when stimulated with a fast enough pulsed signal, numerical trajectories calculated with the different mathematical models start skipping pulses Fig. (2.18). How and how much they skip pulses would ultimately determine the signature of the model.

Once a broad set (20,000) of random PS and IC is generated, a second program reads its output and performs a pulsed stimulation experiment over each of the 20,000 conditions Fig. (2.19).

Each experiment consists of $N_{period} = 20$ different pulsed stimulation sub-experiments (i), each of them with a different input period T_{input} given by eq.(2.3).




```
Nsets = 0
while ( Nsets < 20,000 )

% Random sets
% Generate random set parameters around original ones
for  i=1:Nparam
   generate random number R1 %uniformly dist. [0,1]
   rand_param(i) = original_param(i)*R1*5
end for

% Generate random init. cond. (IC) around original ones
for  j=1:Nspecies
   generate random number R2 %uniformly dist. [0,1]
   rand_IC(i) = original_IC(j)*R2*5
end for

% Solve ODE: cc = calcium, t = time
[ cc,t ] = runge_kutta(rand_param,rand_IC)

% Analyze trajectory cc,t: look for spikes
accepted = false
Nspikes = 0
for i=1:length(cc) - 8
      %spike found
      if cc(i+4) > all four points before and after cc(i+4)

         Nspikes=Nspikes+1
         Tspike(Nspikes) = t(i+4)

         %conditions accepted: 10 spikes found
         if Nspikes == 10
            accepted=true
            Nsets = Nsets + 1
         end if

      end if

end for

% save results
if accepted == true
   Tnat = mean( diff(Tspikes) ) %average spiking period
   save rand_IC, rand_param, Tnat
end if

end while
```

Figure 2.17.: Flux diagram for the random parameter search: each parameter and initial condition is randomly generated following a uniform distribution between 0 and 5 times its original value. The ODE system is solved for each parameter set and spikes are located in the trajectory: a point in the trajectory is considered a spike if it is higher than its 8 surrounding neighbors. Every time a set which allows for oscillations is found (at least 10 spikes are found), its average natural period T_{nat} is calculated. The loop is repeated until 20,000 sets of random parameters and initial conditions are generated for each model.

$$T_{input}(pw, N_{period}, i) = \frac{pw \cdot N_{period}}{i} \tag{2.3}$$

Using the comparison between pulse widths and natural periods for *in vivo* experiments given by eqs. (2.1, 2.2), we fix the pulse width (pw) for these *in silico* experiments at $1/4 \cdot T_{nat}$ with $L = 0.25$. This pulse with lies within the range of the experiments ($0.29 > L > 0.04$), allowing for a proper comparison between the models behavior and the response of real cells. In addition, simulations performed for pulse widths of $1/8 \cdot T_{nat}$ -$L = 0.125$- provided almost identical results. Each of these sub-experiments is performed during $N_{pulse} = 30$ input pulses.

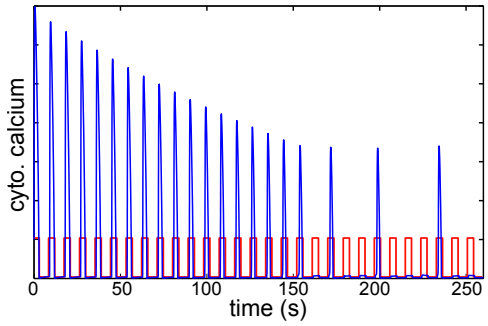


Figure 2.18.: Typical skipping pattern in a model: calcium spikes (blue) occur always when an input pulse arrives (red). For fast enough signals, some pulses are skipped, with little calcium response.

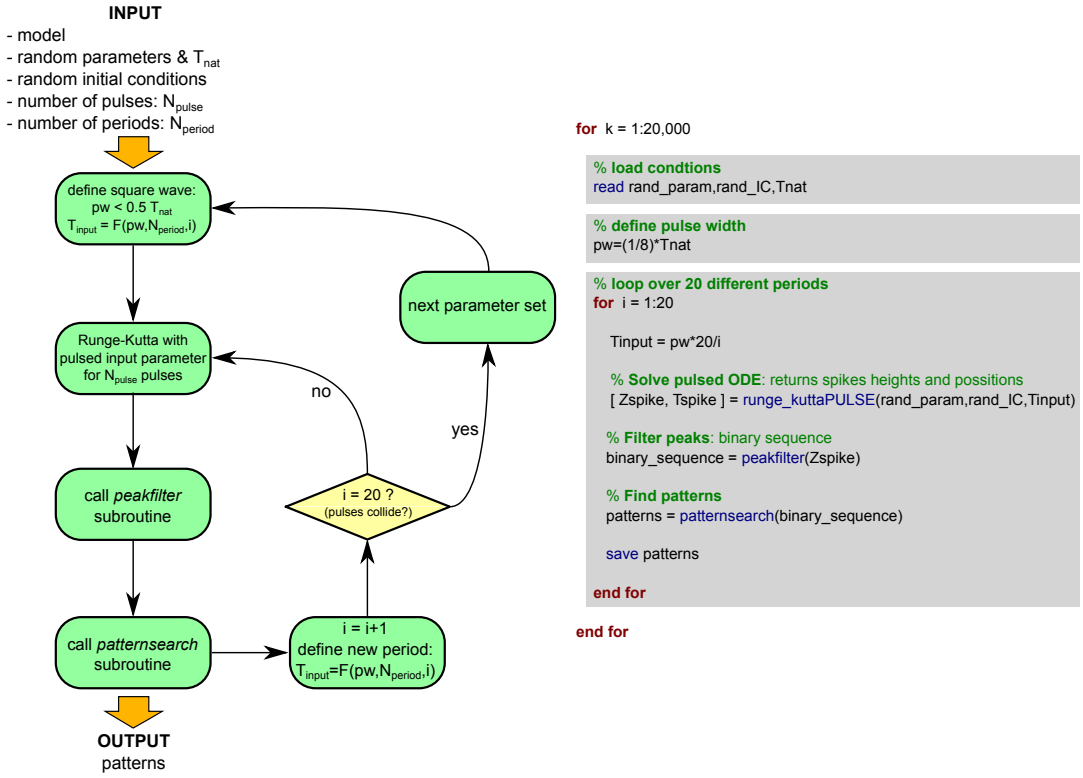


Figure 2.19.: Flux diagram for pulsed stimulation analysis: each set of previously generated random parameters and initial conditions, together with their associated natural period T_{nat} -see Fig.(2.17)-, are red by the program. For each set, N_{period} different input periods are tested. For each period, the system is stimulated with N_{pulse} pulses. For each set of random parameters, T_{nat} is used to fix the time scale: pulse width (pw) is $1/8 \cdot T_{nat}$ and the input periods T_{input} are defined by eq. (2.3) For all the models, we have chosen $N_{period} = 20$ and $N_{pulse} = 30$.

To improve computation time, every time a sub-experiment is performed, the resulting trajectory is analyzed with the *peakfilter* and *patternsearch* subroutines (see next section), as a the first step to generate the model signature.

Pulsed stimulation is the most computationally expensive calculation, and is again written in *Fortran*.

Defining signatures: statistical analysis

Trajectories extracted from both experiments and model simulations are analyzed using exactly the same procedure to ultimately generate the desired signature:

0.- Select all the peaks in the trajectory.

1.- Filter small peaks, which correspond to skipped spikes. From here, a binary sequence is generated.

2.- Look for patterns in the resulting binary sequence.

3.- Generate a signature by doing a histogram of all the patterns.

Once the peaks of the trajectory are found [101], the first step consists on filtering smaller peaks, determining which input pulses have given rise to a calcium pulse and which ones have been 'skipped'. This task is performed with the *peakfilter* subroutine, which filters peaks which are bellow the average height of their 4 closest neighbors, and ultimately generates a binary sequence where 1's and 0's respectively state for accepted and filtered peaks Fig. (2.20).

Second, once the binary sequence is generated it is analyzed in the search for patterns, as indicated in Fig. (2.21). The criteria we use to define 'patterns' gives a good method to quantify the skipping behavior exhibited by each trajectory: we first find the [1,0] sequences, what we have called 'promoters', which indicate the beginning of a pattern (different from the '1' pattern). We then analyze the patterns found in between 'promoters', classifying them as pattern 'N_1/N', where N_1 is the total number of 1's found in between 'promoters' and N the total number of ones and zeros in that specific region. The final output of the program consists of a list of all the patterns found for every sub-experiment (different frequencies) within every experiment (different sets of random conditions). The typical number of patterns found for the whole 20,000 conditions analyzed is of the order of 10^6, giving us confidence on the following statistical analysis.

Finally, for each model a histogram of all the patterns found for every frequency and pulse width is generated. This histogram represents the ultimate signature for each model. The very same procedure is performed for the experimental data, where the patterns found in all the experiments are merged in a single histogram representing the experimental signature. The resulting signatures are plotted in Fig. (2.23).

In order to improve computation time and reduce the amount of data stored, for the model analysis the first three steps (peak finding, peak filtering and pattern recognition) are performed during the simulation itself, using versions of the *peakfilter* and *patternsearch* subroutines written in *Fortran*. On the other hand, once the cell acceptance-rejection filter is run, experimental data is analyzed using the very same *peakfilter* and *patternsearch* subroutines

written in *Matlab*.

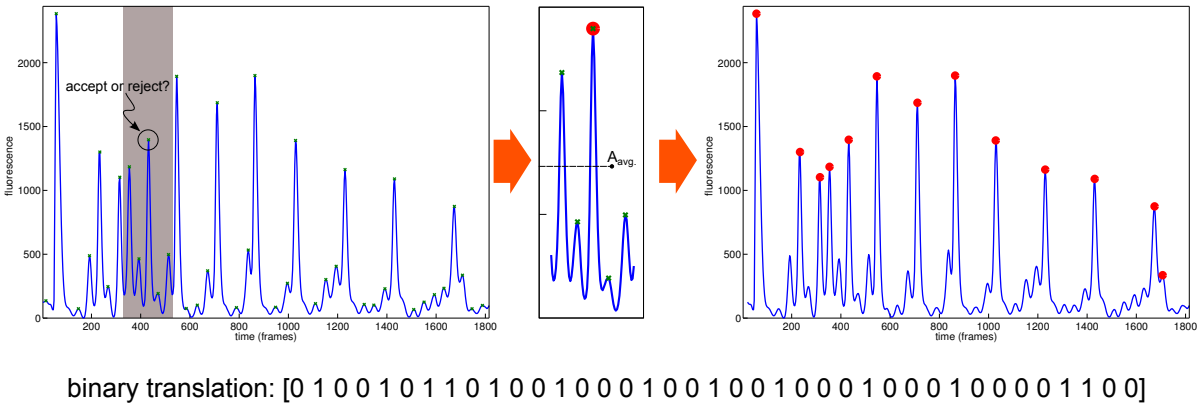


Figure 2.20.: Peak filtering cartoon: once all the peaks are found (green), each peak is compared with its 4 closest neighbors: if it is higher than the average of this local group $A_{avg.}$ it is accepted (red). In order to deal with big spikes, if the standard deviation of the local group is bigger than 0.4 times the local mean, the peak is always rejected. Finally, the overall acceptance-rejection pattern is translated into a binary sequence.

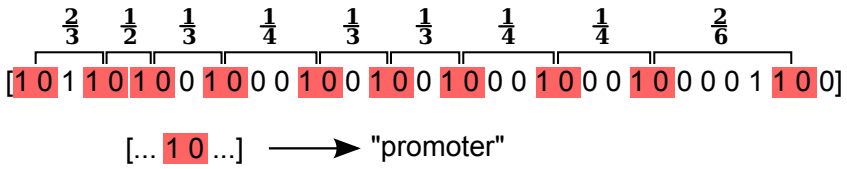


Figure 2.21.: Pattern search method: The [1,0] 'promoter' sequence (red) indicates the beginning and the end of each pattern, which is embedded within successive 'promoters'. Each pattern is named as the number of ones divided by the total number of ones and zeros in a given pattern.

peakfilter

```
- input = Zspike (spike heights)
- output = binary_sequence

for  k = 1:length(Zspike)

   % Define local spikes: four closest neighbours
   mina = max(k-2,1)
   maxa = min(k+2,length(Zspike))
   ind = mina:maxa %index of local spikes

   % Calculate mean and standard deviation of local
   meanlocal = mean(Zspike(ind))
   stdlocal = std(Zspike(ind))

   % Accept or reject spikes: binary sequence
   if  Zspike(k) < meanlocal &  stdlocal > 0.4*meanlocal

      binary_sequence(k)=0

   else

      binary_sequence(k)=1

   end if

end for
```

patternsearch

```
- input: binary_sequence
- output = list of patterns

Npatterns=0

% look for [1,0] "promoters"
for i = 1:length(binary_sequence)-1
   win = [ binary_sequence(i),binary_sequence(i+1) ]

   if  win == [1,0]
      k = k+1
      promoter(k) = i %promoter possitions
   end if

end for

% Analyze patterns within promoters
maxlength=6 %maximum pattern length allowed
for  i=1:length(promoter)-1

   index = promoter(i)+1:promoter(i+1)
   sub_sequence = binary_sequence(index)

   Nones=0
   if  length(sub_sequence) <= maxlength

      for  j = 1:length(sub_sequence)

         %count ones
         if  sub_sequence(j) == 1
            Nones = Nones+1
         end if

      end for

   end if

   Npatterns = Npatterns+1
   pattern(Npatterns) = Nones/length(sub_sequence)

end for
```

Figure 2.22.: Peak filtering and pattern search subroutines

2.4. Results

In the previous sections we have described the context we are working in and the methodology applied to our research: we have introduced the basic features of the IP_3-mediated calcium pathway and explained a novel approach to get a good mathematical description of that system.

Now we present the results obtained when both experimental and computational data are compared. The results are presented in two stages: in the first stage we describe the outcome of the 'cellular interrogation' routine, showing how some models give a good description of the experiments and are accepted while others are rejected. In a latter stage the remaining models are analyzed in detail, extracting from them information about the molecular machinery they describe and making predictions for further experiments.

2.4.1. STAGE 1: model acceptance-rejection

First test: comparing signatures

Once all the different signatures are computed, we can compare them with the experimental one and start discarding models. Signatures are plotted in Fig. (2.23):

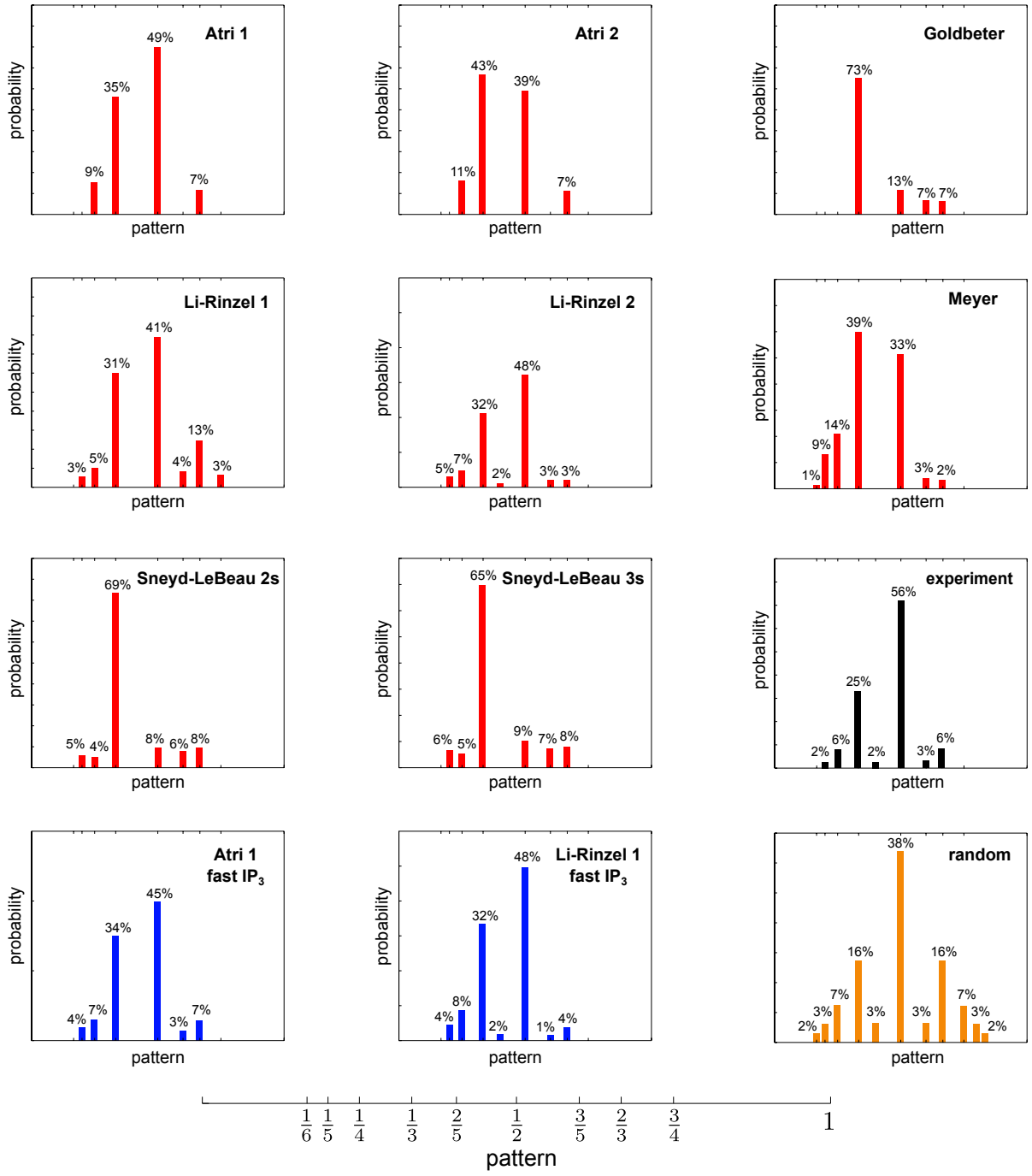


Figure 2.23.: Signatures of the original models (red), modified models (blue), a random distribution (orange), and the experiments (black).

At first glance we can easily notice that signatures from Goldbeter, Sneyd-LeBeau2 and Sneyd-LeBeau3 are completely different form the one generated by real cells. At the same time, it seems like Atri2 and Meyer's signatures also differ from the experimental one: while the experimental distribution shows the higher bin at $1/2$ with bins in the $1/n$ positions decaying with n, this two models have distributions with the $1/3$ pattern being the most abundant one. On the other hand, it seems like Atri1, Li-Rinzel1 and Li-Rinzel2 fit the observed distribution.

To compare the models and experimental distributions in a more quantitative way, for each of them we compute their χ^2 as:

$$\chi^2 = \sum_{bins} \frac{(H_{bin} - E_{bin})^2}{E_{bin}} \tag{2.4}$$

Here H_{bin} and E_{bin} respectively state for the height of each bin in the models and experimental distributions. Fig. (2.24) shows the values of χ^2 for all the models, in addition to its value computed for the result of looking for patterns in a random distribution of ones and zeros.

The more similar the model and experimental signatures, the smaller the computed χ^2. Following this criteria we easily discard Golbeter and Sneyd-LeBeau (SLB) models. Although Atri2 and Meyer models have values of χ^2 which are considerably smaller, they are still more than twice the values of Atri1, Li-Rinzel1 and Li-Rinzel2. Therefore, we discard Atri2 and Meyer too, keeping Atri1, Li-Rinzel1 and Li-Rinzel2.

To finish, notice that the signature of a random series of ones and zeros gives a relatively low χ^2. This is given by the fact that finding small patterns like a $1/2$ pattern ($[1,0,1]$) is more likely than finding longer ones like a $1/3$ ($[1,0,0,1]$). Therefore, the overall random signature ends up having its higher bin at $1/2$ with bins decaying symmetrically around it. Although this random signature is similar to the experimental one for patterns below $1/2$, this similarity is just a coincidence: when we look at the skipping patterns presented by real cells or by the selected mathematical models, we see that this skipping behavior is by no means random. The skipping process always follows the same order: at first spikes occur for each input pulse, then we find a region in which one spike occurs every two pulses, followed by regions where two pulses are skipped, then three and so on. Different cells start the skipping process at different stages, but they always follow the described sequence. Moreover, although the computed χ^2 for a random series is relatively high, it is still well above Atri1, Li-Rinzel1 and Li-Rinzel2's value, mainly because of the lack of bins above $1/2$ in both models and experimental distributions.

It seems complicated to give a formal justification to why some models have one signature or another. The fact that some spikes are skipped when the system is stimulated with a pulsed input seems easy to understand:

The input of the system is nothing more than a parameter (R) of the system whose value

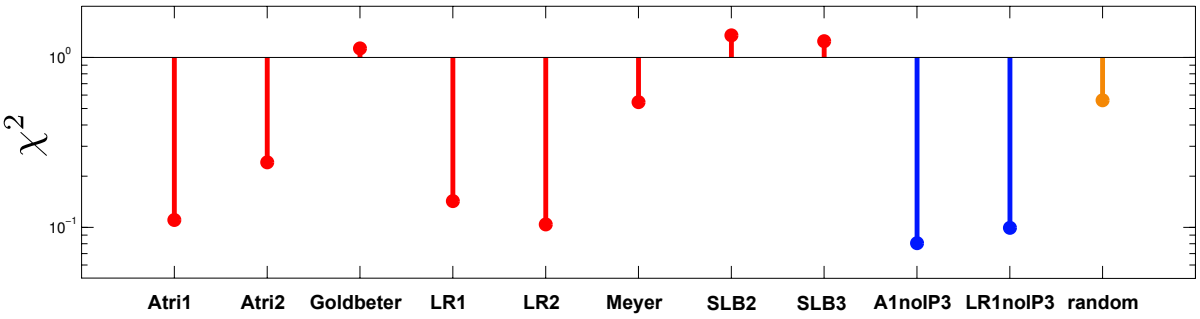


Figure 2.24.: χ^2 computed for all the models signatures, in addition to a distribution computed for a random series of ones and zeros: the lower χ^2 the closest the model signature is to the experimental one. Here SLB states for Sneyd-LeBeau models.

oscillates with time. When this parameter is set to its 'normal' value, R_{max}, the system has a limit cycle and oscillates around it, while the cycle disappears when R is set to 0. Then, if this parameter is alternatively changed from 0 to R_{max}, generating a square wave, the limit cycle would alternatively appear and disappear. Depending on the state of the system at which this transition occurs, it may complete an extrusion or rapidly fall to its steady state Fig. (2.25). Within this picture, the faster the input is turned on and off, the more likely would be to find the system just at the beginning of the extrusion, leading to what we have called a skipped pulse. Therefore, the faster the input signal, the more pulses are skipped.

On the other hand, the exact abundance of each pattern when random values are assigned to the parameters of the system is something that cannot be extracted from the dynamical models without running the simulation. Nevertheless, although we cannot say that this specific feature of this model gives rise to more abundance of $1/3$'s instead of $1/2$'s and so on, we expect that the overall structure of the model is ultimately responsible for this behavior. The next arguments will show how this expectations seem to be, at least in part, correct. Performing more tests over the accepted models, we are able to reproduce the typical experimental behavior with 2 models out of three.

Second test: analyzing spike decay

So far we have ruled out five models out of the eight originally proposed. Having a look at the three remaining models we notice that one of them corresponds to a class 2 model (with feedback from cytosolic calcium to IP_3) while the other two are class 1 models (with no feedback) Fig. (2.15). The fact that this classification of the models is based on a fundamental property like the existence the calcium-IP_3 feedback, indicates that further tests should be conducted in order to discard one class or another: in principle we could accept to find various models giving a good description of the system, but only as long as they give relatively similar descriptions. Otherwise, our method would be useless, being unable to resolve even the most basic interactions.

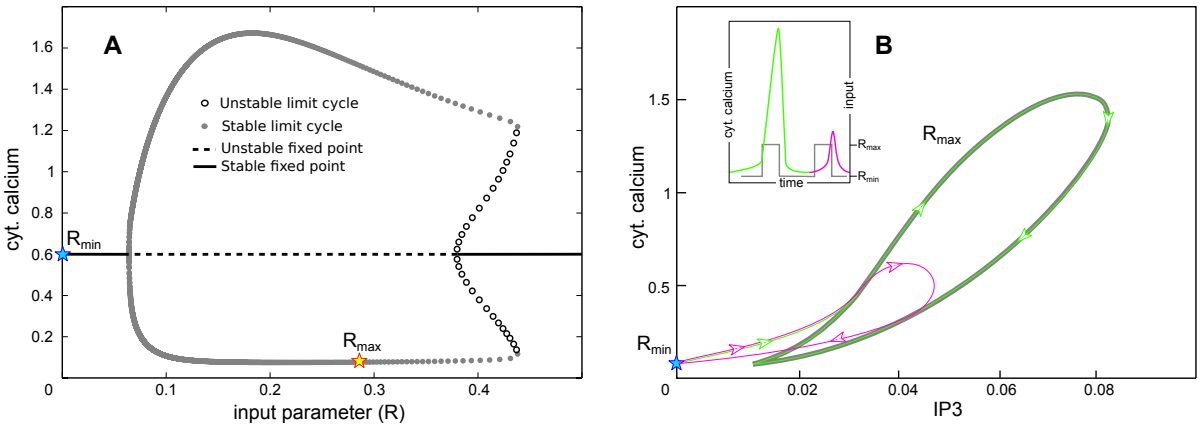


Figure 2.25.: Bifurcation analysis of the pulsed stimulation: the bifurcation diagram (A) shows a Hopf bifurcation, in which a stable fixed point (filled line) leads to an unstable fixed point (dashed line) and a stable limit cycle (gray) upon a change in the value of the input parameter R. A square wave input alternatively changes the value of R from 0 to 0.28, moving the system from one regime to another. In the phase diagram (B) it is shown how trajectories exhibit bigger (green) or smaller (pink) calcium spikes depending on the moment the transition between R=0 and R=0.28 occurs.

To complete our 'cellular interrogation' protocol, a second test should be developed. The easiest way seems to be to try and reproduce the experimentally observed calcium trajectories using the selected models.

When stimulated with a constant histamine concentration, most of the cells exhibit trajectories in which the amplitude of calcium spikes decay smoothly Fig. (2.9). This behavior could be a product of receptor desensitization [102], or be produced by ER calcium depletion. While the first case is not included in any of the models, the latter can be easily tested by fine tuning the initial conditions. If we start the stimulation with the ER calcium concentration set above its maximum value in the limit cycle, we find out that the system would follow a spiral until the steady state orbit is reached. This is translated in calcium spikes of decaying amplitude. While the system is spiraling down into the limit cycle, the ER calcium decreases until the steady state is reached Fig. (2.26).

How fast this spiral falls down into the limit cycle would be determined not only by the specific parameters but also by the structure of the system itself Fig. (2.27). We observe that Li-Rinzel1 and Atri1 show spirals which take more than an orbit to fall into the limit cycle, exhibiting calcium spikes whose amplitude decays smoothly until a steady state is reached. On the other hand, spirals in the Li-Rinzel2 model always reach the steady state limit cycle in less than an orbit, giving rise to an abrupt decay of the spikes amplitude. This fast amplitude decay is the result of the speed-up in the adaptation time produced by a positive feedback in the system: the higher the calcium concentration the higher the IP_3 production rate and therefore the faster the ER depletion.

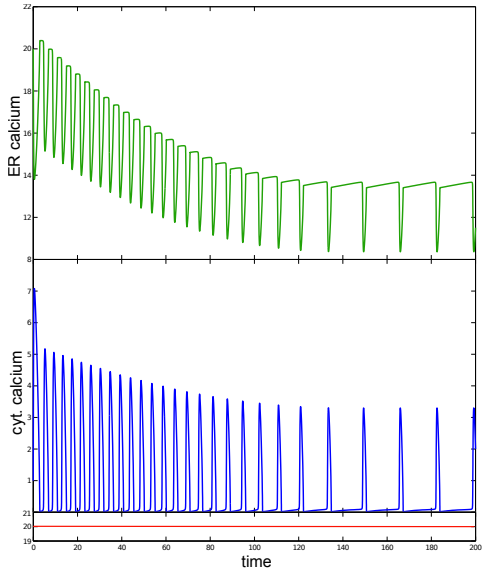


Figure 2.26.: Class 1 Atri model trajectory: starting with a high ER calcium concentration (green), a constant input (red) is applied. ER calcium concentration decays until the system reaches its steady state. Meanwhile, cytosolic calcium peaks (blue) experience an amplitude decay.

This behavior is confirmed under parameter variation Fig.(2.28). Different sets of random parameters give rise to different time-scales in spike decay. However, regardless the time-scales, Atri1 and Li-Rinzel1 always show a smooth decay, while this decay is always abrupt in Li-Rinzel2. To quantify this property, we analyze how many spikes are needed to fall half the way between the maximum and the minimum spikes for all three models under different random sets of parameters. As expected, we get that Li-Rinzel2 almost always needs only 1 spike to get to the steady state, while Atri1 and Li-Rinzel1 spend multiple spikes to reach that situation. In addition, this analysis is applied to real cells stimulated with a constant histamine concentration, showing that an average decay time of around 4 spikes.

Since only Li-Rinzel1 and Atri1 models are able to reproduce the experimental behavior for constant stimulation, this allows us to easily discard Li-Rinzel2.

It is important to notice that we have directly assumed that ER depletion is responsible for the decaying amplitude behavior which gives us the possibility to discard Li-Rinzel2. Although at this point this may seem arbitrary, in the next section we will perform more tests on the two remaining models which will give us more confidence about their abilities to describe the system, making it clear that they provide a very good description of the IP_3-mediated calcium pathway in Hela cells under histamine stimulation.

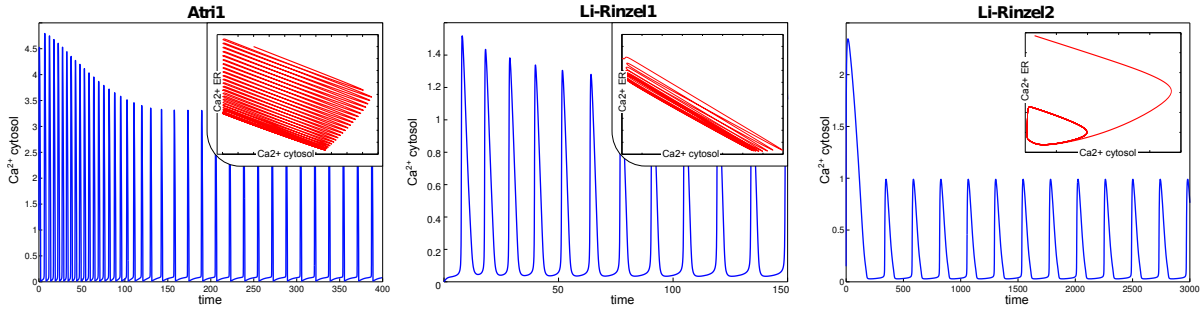


Figure 2.27.: Trajectories for a constant input (blue): class 1 Atri and Li-Rinzel models exhibit smooth decaying amplitude when an initial high ER calcium concentration is set. On the other hand, class 2 Li-Rinzel model never shows such behavior. Phase space representations (red) show how trajectories in class 1 models complete various orbits before falling down to the limit cycle when the latter initial conditions are set, while class 2 Li-Rinzel model never completes a whole orbit.

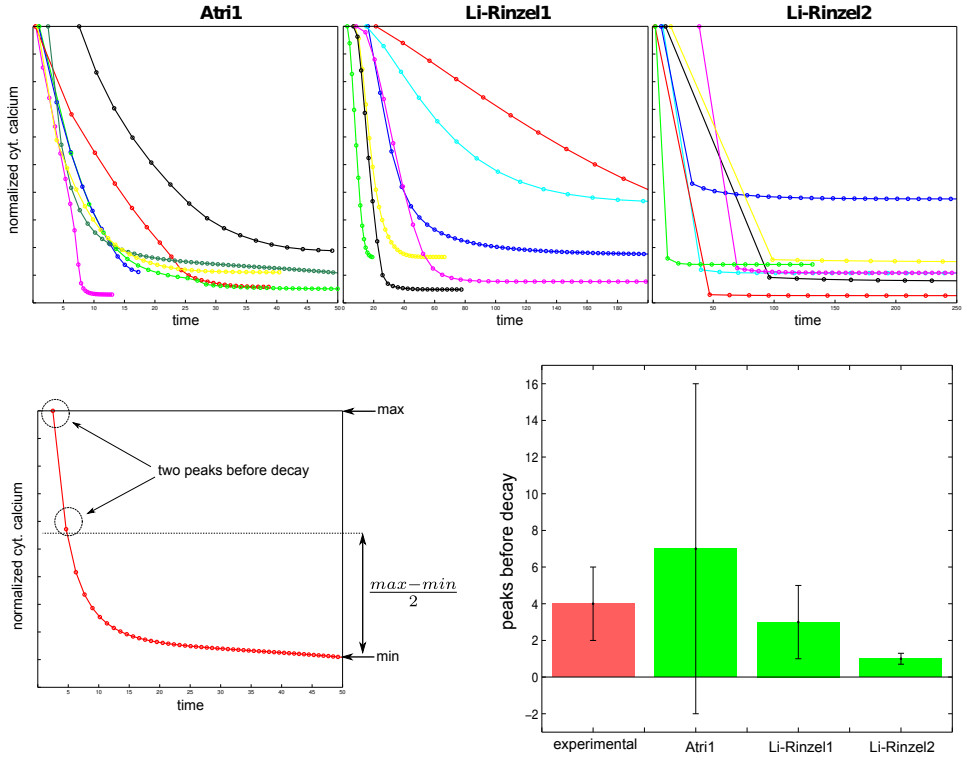


Figure 2.28.: Envelope trajectories of peak positions for different random conditions in which peak decay is observed. In real cells, calcium spikes decay in around 4 cycles. Atri1 needs an average of 7 cycles, while in Li-Rinzel1 this average is around 3. On the contrary, Li-Rinzel2 almost always needs only 1 spike to fall into the limit cycle.

2.4.2. STAGE 2: model refinement & prediction making

Stage 1 allowed us to find the best two candidates to modelize our system. Unique model discrimination is not always possible when working with complex networks; however, we have

been able to select two models which account for the very same interactions and features of the system, although described with slightly different dynamics.

They are both class 1 models, with no feedback from cytosolic calcium to IP_3 and therefore, no IP_3 oscillations occurring during spiking. Unlike Sneyd-LeBeau models, they do not rely on a specific description of the IP_3r's to achieve oscillations and unlike for Goldbeter model, a negative feedback exists between cytosolic calcium and the amount of calcium that IP_3r's release into the cytosol. The biggest difference we find is that Li-Rinzel1 includes a term which accounts for a calcium flux (J_{noIP3}) form the ER to the cytosol which is independent of the IP_3 concentration (see equations in appendix A). In principle, this extra flux term does not seem relevant for the overall behavior of the system, since it just changes the ER and cytosol calcium concentrations when no IP_3 is present, without producing oscillations

In the rest of the *Results* section we will show how these models are able to reproduce not only the experimental behavior for a constant stimulation, but also the one observed for pulsed stimuli. This will lead us to propose a modification of the models in which IP_3 dynamics is supposed to be much faster than the rest of the variables of the system. Under this new premise, both models reproduce the experimental behavior even better, leading to signatures which are even more similar to the experimental one.

Modified class 1 Atri and Li-Rinzel models

Class 1 Atri and Li-Rinzel models can deal with the experimentally observed peak adaptation when a constant input is applied. When we start to work with pulsed inputs, the models do not exactly behave like real cells.

Real cells always spike entrained with the input: spikes only occur when an input pulse arrives. Therefore, there is no inertia on the input, meaning that the output almost instantaneously reacts to it. This suggests that all the molecular machinery involved in propagating the external signal from the cellular membrane (PLC, IP_3...) to the calcium-oscillations machinery (ER, IP_3, IP_3r...) has a fast response time: as soon as the input pulse is switched off, this machinery responds switching off the oscillations, while it suddenly responds when the input is turned on again, turning on the oscillations. This is equivalent as saying that no effective high pass filter effect is found for the input-propagation machinery at the input frequencies we are working, therefore delimiting the possible time-scales involved in IP_3 dynamics.

When trying to reproduce with Atri1 and Li-Rinzel1 models the behavior experimentally found for square wave stimulation, we find out that, for fast enough pulses, there is a loss of entrainment between the input and the output Fig. (2.29). This loss of entrainment is produced by IP_3 dynamics: when the input is turned off, the system keeps spiking until IP_3 is degraded under a threshold value Fig. (2.32).

Therefore, Atri1 and Li-Rinzel1 do not have a fast input-propagation machinery. Both

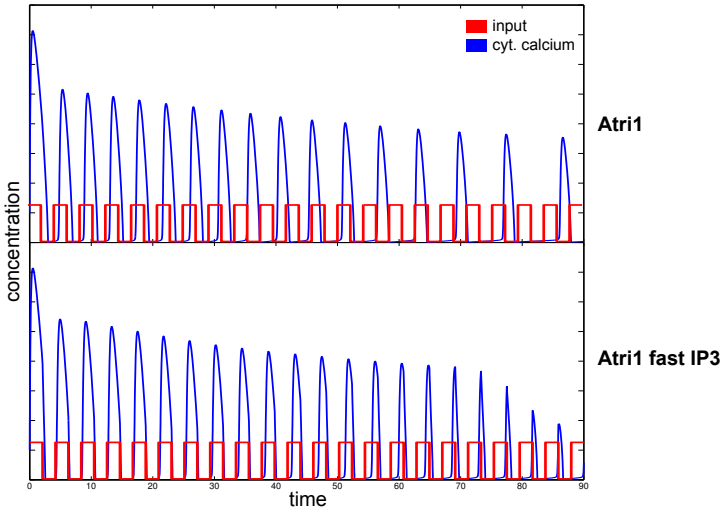


Figure 2.29.: Entrainment loss in Atri1: IP_3 inertia implies that, when the input period is small, IP_3 does not have enough time to react, and therefore, the input state is not well transmitted to the spiking machinery. This is translated into a loss of entrainment between input and output, where spikes are found in between pulses. On the other hand, when fast IP_3 dynamics is implemented, signals from the input to the spiking machinery are perfectly transmitted, and therefore spikes are only found synchronized with the input pulses.

models follow a diagram similar to Fig. (2.30). The input acts directly controlling IP_3 steady state concentration p_{st}, which ultimately is responsible for the spiking machinery activation. IP_3 dynamics follows

$$\frac{dp}{dt} = ir(p_{st} - p) \tag{2.5}$$

Here, IP_3 has a characteristic time of $\frac{1}{ir}$, which ultimately slows down signal propagation.

Since these models do not show IP_3 oscillations following calcium spikes, IP_3 dynamics is not responsible for the oscillatory machinery at all. IP_3 is just a second messenger which carries the external input information to the independent oscillatory machinery. This allows us to refine the signal propagation capability of both models by supposing fast IP_3 dynamics. Thus, we neglect IP_3 dynamics by assuming that its response-time is fast enough to respond to the input instantaneously. Since IP_3 steady state p_{st} was already used as the input for both models, now we have directly $IP_3 = p_{st}$, and IP_3 is supposed to be the input itself. We end up with modified class 1 Atri and Li-Rinzel models which follow the diagram presented in Fig. (2.30). This new version of the models allows us to reproduce the experimental behavior as shown in Fig. (2.31) and Fig(2.33).

To ultimately test the fast IP_3 dynamics hypothesis, we perform a series of experiments in which cells ares stimulated with a constant histamine concentration for 10 minutes, then histamine is removed and, after 10 minutes, histamine is applied again. If IP_3 dynamics is

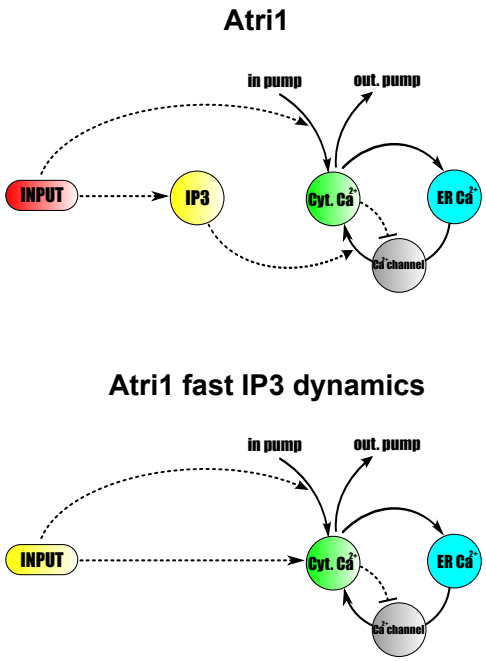


Figure 2.30.: Modified Atri1 schematic: diagram of the original class 1 Atri model and the simplified one (same modification is made on class 1 Li-Rinzel model). In A, the input ($p_s t$) stimulates the production of IP_3. IP_3 binds to the $IP_3 r$ in the ER membrane and opens the calcium channels, leading to the first spike. calcium feedback closes these channels, stopping calcium release until cytosolic calcium concentration is low again and a new spike occurs. Notice that IP_3 is just responsible of input propagation and has no feedback from the oscillatory mechanism. calcium oscillations are produced without IP_3 oscillations involved, since no feedback from calcium to IP_3 is present. In B, all the oscillatory mechanism is the same, but instead of having an input which stimulates IP_3 production, IP_3 is set as the effective input

fast, we would expect cells to stop spiking just after histamine is washed away, and then start spiking as soon as the second histamine pulse arrives. This is in fact what we found, confirming that the fast IP_3 dynamics hypothesis is correct Fig. (2.32). This method allows us to estimate the average cells activation and deactivation times -respectively $T_{act.}$ and $T_{deact.}$-, defined as the time needed by intracellular calcium to start spiking again upon increase of histamine concentration.

Now that we have refined the remaining models in order to be able to reproduce the experimental trajectories when a pulsed input is applied, we can have a look at their signatures, which were already included in the signatures figure Fig.(2.23) and the χ^2 analysis Fig.(2.24). Doing so we find out that, although the new signatures are very similar to the original ones, they actually improve, with a proportion between pattern abundances closer to the experimental one. This improvement could be a more indirect evidence of the difference between the dynamics of IP_3 and the rest of the variables. To illustrate how these modified models are able to reproduce the experimental behavior, trajectories corresponding to pulsed stimu-

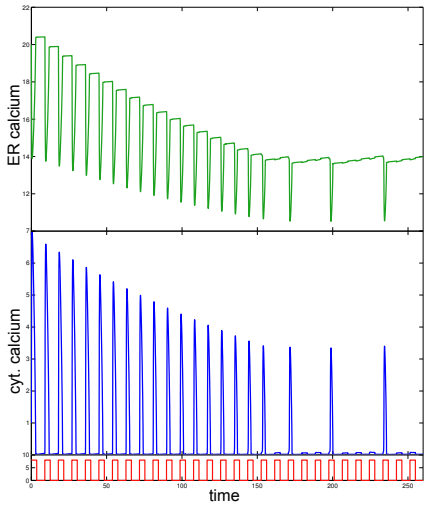


Figure 2.31.: Modified Atri1 model trajectory: similarly to the case presented in Figure(2.26), ER (green) and cytosolic calcium (blue) are represented together with the input (red). Starting with a high ER calcium concentration, the system initially follows the input step by step. Close to the steady state region (around time = 150 a.u.), the system has slowed down and it is not able to follow the driving frequency anymore. Instead, it starts skipping pulses, first one ($\frac{1}{2}$ pattern), then two ($\frac{1}{3}$) and then three ($\frac{1}{4}$). The simulation has been performed using an input frequency which is twice the steady state natural frequency and a bin width of 0.6 the natural period

lation expeirments are compared with the analogous simulations performed in the modified Atri1 and Li-Rinzel1 models Fig.(2.33)

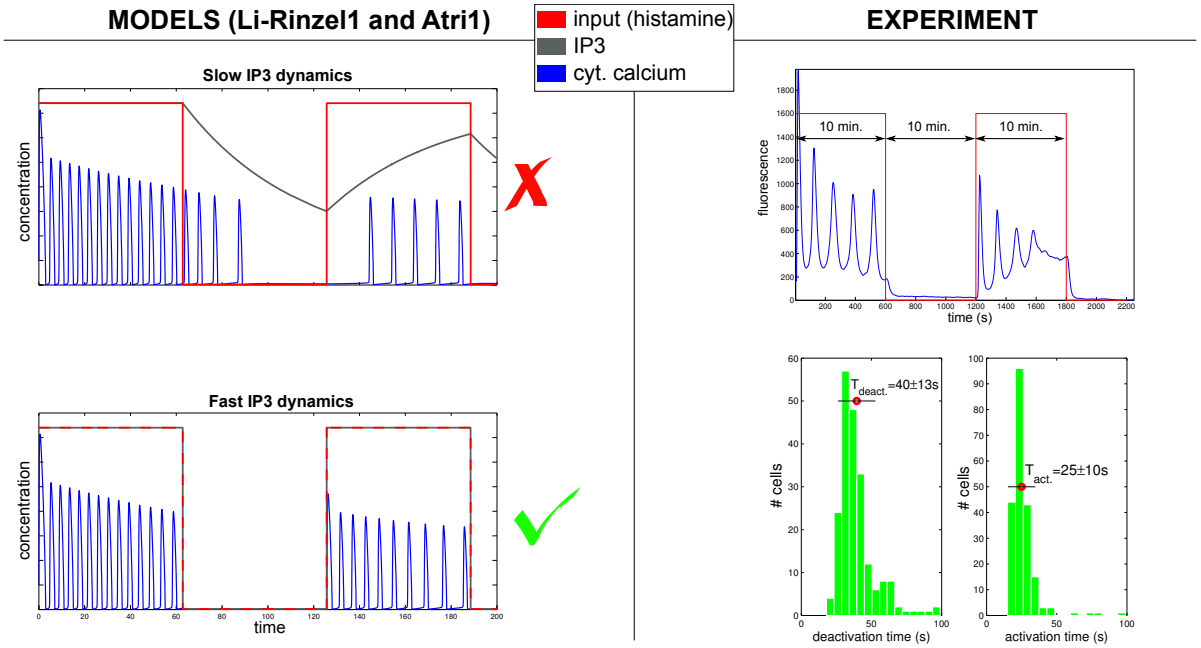


Figure 2.32.: IP_3 decay in Atri1 (similar behavior in Li-Rinzel1): when the input (red, $p_{st} \equiv IP_3$ equilibrium concentration) is switched off, calcium (blue) keeps spiking because IP_3 (grey) needs some time to degrade. When IP_3 reaches a threshold value the system stops spiking. If the input is turned on again, IP_3 spends some time increasing its concentration until it exceeds the threshold and spiking starts again. When fast IP_3 dynamics is implemented, the system reacts immediately to changes in the input. If the input is turned off so does spiking. When it is turned on again spiking starts at the same time (although spikes are smaller due to previous ER calcium depletion). Performing this type of stimulation on real cells we obtained trajectories consistent with the fast IP_3 dynamics scenario. Narrow distributions on activation and deactivation times -respectively $T_{act.}$ and $T_{deact.}$- show that this behavior is exhibited by most cells.

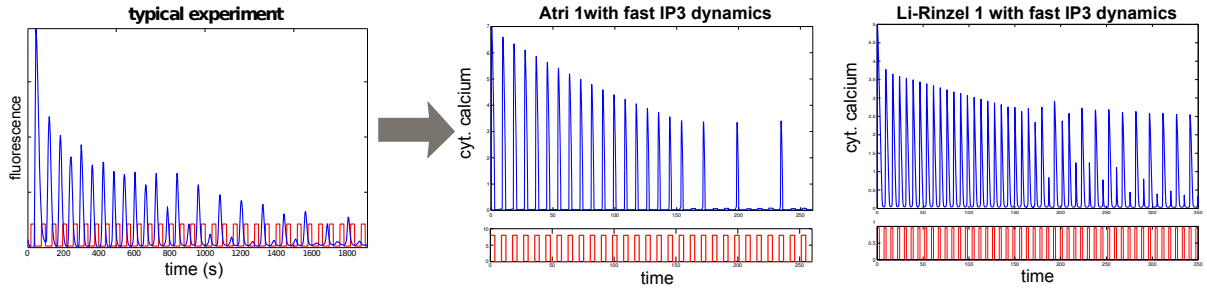


Figure 2.33.: Mimicking pulsed stimulation: comparison of the typical experimental response for pulsed stimulation with the analogous trajectories for modified class 1 Atri and Li-Rinzel models.

Predictions

Now that we feel confident with the selected models and we have refined them to better describe the experimental behavior, they can be used to make some predictions about the real system. To do so, we have written a visual *Matlab* GUIDE program which is basically a virtual microfluidic device. It allows us to load the models in the format given by PROTEUS [103] and stimulate them with different pulsed stimuli, as well as to modify the initial conditions and parameters. We use it to perform *in silico* microfluidic experiments on the modified Atri1 and Li-Rinzel1 models which are useful to propose new *in vivo* experiments. A list of the predictions is provided in Table (2.4).

observed behavior	**cause**
always entrainment	IP_3 dynamics faster than the typical time scale of the oscillations (CONFIRMED)
decaying amplitude	Initial ER calcium concentration is high (above its steady state level)
big initial spike	Initially, most IP_3 receptors are active IP_3 concentration is high (or fast dynamics)

Table 2.4.: Predictions on calcium behavior provided by both class 1 Atri and Li-Rinzel models

The first two predictions have been widely discussed on the previous section: fast IP_3 dynamics is responsible for the entrainment we see in all the cells from all the experiments. This prediction has already been confirmed with a two-step stimulation experiment Fig.(2.32) Although receptor desensitization can not be discarded and may be possibly involved in the process, slow depletion of the initially full ER seems to play a main role in the observed decaying amplitude. Receptor desensitization could be included in the models simply by simulating a decaying IP_3 concentration, as a result of less ligand being detected by the cell.

Another feature we usually find in experiments performed under both constant and pulsed stimulation is an initial big spike Fig. (2.9). Although decaying spikes imply that the first spike is the higher one, this spike is usually higher than expected, taking into account the decay rate of the following spikes. This behavior can be reproduced by both Atri1 and Li-Rinzel1 models when the initial proportion of ready-to-go IP_3 receptors (n) and IP_3 concentration are high Fig. (2.34). The same effect is observed if IP_3 dynamics is fast, since it gets to its steady state "instantaneously".

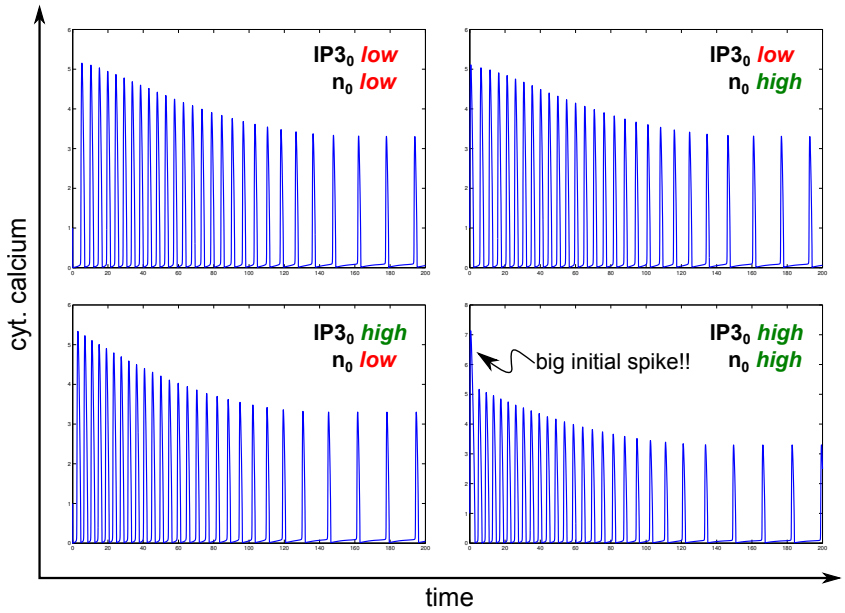


Figure 2.34.: Trajectories under different initial conditions for Atri1: left to right, up to down: low initial IP_3 concentration ($IP3_0$) and fraction of ready receptors (n_0), low IP_3 concentration with high fraction of ready receptors, high IP_3 concentration with low fraction of ready receptors, both high IP_3 concentration and fraction of ready receptors. Observe that the big initial spike is only found for when both IP_3 concentration and the fraction of ready receptors are high.

2.5. Discussion

In the introductory section of this chapter we have emphasized the importance of the IP_3-mediated signaling pathway, which plays a key role in vital cellular processes and whose specific details are still to be understood. At the same time, this system has been introduced as a well studied example of an oscillatory system, being a perfect field of action to develop new techniques suitable for the quantitative characterization of such systems.

Therefore, using this system as a prototype, we have obtained novel results on to fronts: on the one hand, we have managed to find a way to obtain biologically relevant answers for previously unknown molecular mechanisms, while on the other hand we have developed new experimental and computational approaches suitable to perform a systematic study of oscillatory systems in general.

All the experimental work has been done using Hela cells stimulated with histamine, and therefore, given the broad spectra of responses described in the literature depending on both cell type and stimulus, all the results we have obtained are restricted to this type of cells under this specific stimulation.

By following our analysis protocol we have ended up with two similar models which give a good description of the IP_3-mediated pathway: Atri1 and Li-Rinzel1. The fact that this

two models are class 1 models, with no feedback form calcium to IP_3, is consistent with the recent results obtained by Matsu-ura et al. [93], where they monitored IP_3 precisely in Hela cells.

Moreover, it seems likely that the effect of mitochondria (proposed in the Meyer model) is not specially relevant to produce oscillations, although its effect may be mimicked in the calcium leak to the extracellular media (J_{out}) present in both Atri1 and Li-Rinzel1 models.

The only different interaction in both models is the non-IP_3-mediated flux form the ER to the extracellular media present in Li-Rinzel1. Since both models give a good description of the experimental results, we cannot accept or reject the existence of such interaction. When we remove it from the equations of Li-Rinzel1 we basically get the same behavior, so it does not represent an important influence in the overall dynamics of the model. Nevertheless, we may intuitively expect such a flux to be somehow present due to the non-zero probability of an IP_3r to be activated by calcium ions which are always present in the cytosol.

On the contrary, the dynamics of the IP_3r and the way they release calcium seems to be the key mechanism governing the final behavior of the model. Working with PROTEUS [103], we have performed a mechanistic study in which hybrid models are created by replacing different mechanisms of Atri1 (*leaks*,IP_3r's...) by the description given by other models. Preliminary results show that Atri's IP_3r dynamics plays a fundamental role in the ability of the model to exhibit the experimentally observed smooth decaying spikes and increasing inter-spike period.

On the experimental side, we have developed microfluidic devices which could be in principle used to study many different cell lineages. Moreover, they allow us to stimulate cells with different ligands, which can be applied to create a broad set of temporal patterns. Once the building protocol has been established, their fabrication is relatively straight forward, being able to produce around 18 devices at a time and therefore allowing to perform long series of experiments.

The main characters in the 'cellular interrogation' protocol are the 'signatures'. The way we have defined a signature is somewhat empirical, but by trial and error we are able to give a good operational definition. Looking back at the experimental signature, which has been computed doing statistics over all the experiments together, one could argue that we obtained that specific signature by chance, that repeating the experiments would give rise to a completely different histogram. That is not the case: when we compute individual signatures for each experiment we end up obtaining basically the same pattern for all of them, confirming that this pattern is recurrently found within many experiments.

When applying our signature definition to different mathematical models we end up selecting three models, one class 2 model and two class 1. Further analysis helps us to discriminate between these models: while amplitude decay in the class 2 model occurs fast due to the calcium-IP_3 feedback, this decay is much slower in class 1 models. Comparing this am-

plitude decay with the typical experimental decay we discard the class 2 model, concluding that no calcium feedback on IP_3 dynamics is present. The fact that the signatures exhibited by the remaining class 1 Atri and Li-Rinzel models match that of the experiments remarks the consistence of the overall protocol.

To conclude, we could say that we have developed a method which, at the end of the day, is purely experimental: on the one hand we have our microfluidic experiments which obtain data from real cells subjected to complex stimulus, and on the other we have our computational experiments in which data is obtained from the behavior of previously published mathematical models. We do not focus on the analytical details of the models until we end with the whole protocol, in which the models whose responses better matches the experiments are selected. At this point, we start having a look at the chosen models to infer properties of the real system and make predictions for future experiments. Among these predictions, we are able to experimentally confirm the fast IP_3 dynamics hypothesis. A workflow diagram of the whole process is provided in appendix A, Fig. (A.1).

3. Signal detection and propagation: gene networks

In the final chapter of this thesis we will focus on some properties of the central decision-making machinery inside cells. We will analyze the signal detection and propagation properties of simple gene network motifs, which can be in some extent applied to larger networks. First, we will show how different network topologies exhibit different trade-offs in their signal detection and noise filtering capabilities. To conclude, the links between the dynamic response, the structure, and the signal propagation abilities of the circuits will be addressed.

3.1. Introduction

Most transduced signals ultimately modify the gene expression level of the organism, activating and repressing the synthesis of key proteins which contribute to trigger the appropriate cellular responses. As we already mentioned in the introductory chapter, depending on the nature of the original stimuli, transduced signals can exhibit a wide range of spatio-temporal characteristics, ranging from steady increases to oscillations or spikes in concentration of specific transcription factors. At the same time, these signals are more or less noisy, mainly due to the low number of molecules involved.

Gene regulatory networks detect and process transduced signals by means of circuits of interacting genes. The topology of these circuits is directly related to the type of signals they have to process and the responses they generate: some structures favor the detection of fast oscillatory stimuli, while others only respond to sustained changes in signal concentration. Moreover, topology influences the networks robustness against noise, which is more or less important depending on the context they are operating in. For instance, oscillations of transcription factor $NF-\kappa B$ are activated upon cellular stress and immune response, determining timing and specificity in the activation of downstream genes [104]. Mathematical modeling of this system helps to prove that negative feedback loops are responsible for the circuit behavior. On the other hand, feedforward loop interactions seem to be responsible for responses to persistent stimuli in *Escherichia coli* transcription networks [105].

Gene networks are huge, formed by thousands of interacting genes. For example, the genome of *Escherichia coli* is known to be comprised of around 4,200 genes, while budding yeast *Saccharomyces cerevisiae* has around 6,000 genes. More complicated organisms have even

larger networks, although apparently more complexity does not directly imply a bigger genome: *Caenorhabditis elegans* has around 19,000 genes, *Homo sapiens* around 23,000 and corn *Zea mays* around 30,000 [5]. How can we develop a theoretical framework to study such complex networks in a methodical way?

A possible approach to understand the complex functioning of regulatory networks is to dissect them in simpler modules, whose function can be individually characterized. Within this description, each module is assumed to work more or less independently, processing external signals and triggering responses while passing its output result to the rest of the network. The combined action of many different modules with specific signal-processing characteristics gives rise to a vital coordinated response. For instance, coordination of executor proteins governing cell cycle regulation in yeast is regulated by coupled sub-circuits with diverse signal processing characteristics [21].

The structure of gene networks is not only modular, Alon a co-workers found that, among the modules forming those networks, some structures -network motifs- were significantly abundant compared to what would be expected in networks of random topology [18]. Why are these specific structures overrepresented all around the genome, even for evolutionary distant organisms like bacteria or yeast?

The study of the signal processing capabilities of simple genetic modules is of key importance in two distinct fronts: on the one hand, it helps unveiling the reasons which make network motifs so important for the overall function of the genome [19–21]. On the other hand, it contributes to create a toolkit which can be used to engineer synthetic gene circuits able to perform specific tasks [106, 107].

In spite of all the previous works, so far no general framework has been presented to methodically characterize all the signal-response characteristics of simple gene networks. Here we present a first attempt to do so: we have developed a formalism which allows us to make predictions of signal propagation properties, noise, and response characteristics of any three component module as a function of its topology i.e. interactions between genes and their strengths.

In the first part of this chapter we present the results obtained in the study of signal detection and noise filtering capabilities, showing the different trade-offs exhibited by diverse topologies and the diverse regimes in which different circuits are able to operate.

Intuitively, it seems clear that the structure of the circuit would be responsible not only of its signal detection properties, but also of the speed of its response to fast changes in input concentration, what is commonly known as response times. Inspired by this fact, in the second part of the chapter we show how these three features are related.

3.2. Methods

Here we present the context in which all the study takes place. We first provide a description of the simple three component module we will use to modelize simple gene networks, followed by a description of how the strengths of the interactions are quantified. To finish, we present the theoretical framework used to analyze signal propagation and dynamical characteristics of those modules.

3.2.1. A model for a three component module

A general three component module provides us with a simple enough yet complete model to study basic gene interactions Fig.(3.1). The module consists of an input species I which positively interacts with the sensor component S, enhancing its production. Depending on the case, the sensor component acts as an activator or a repressor of the output species O. Additional interactions may be present: feedforward loop interactions, in which the input directly activates or represses the output, both positive and negative feedback loop interactions going from the output to the sensor component, and autoregulations, in which the sensor and/or the output components enhance or repress their own production. The strength of the interaction between each pair of species ij is given by the its susceptibility s_{ij}, whose analytical form is provided in the next subsection.

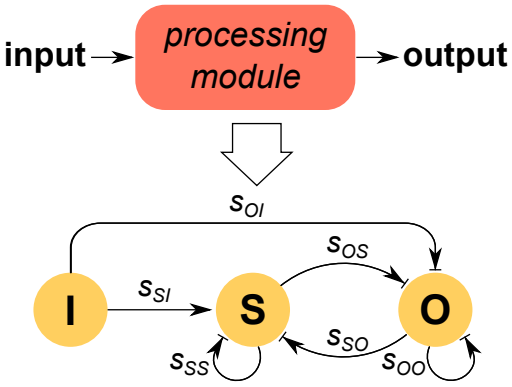


Figure 3.1.: Schematic of a general 3 component module: an input species (I) activates the production of the sensor component (S). The output (O) production is either positively or negatively regulated by the sensor component. In addition, we could have a feedforward interaction between the input and the output or a feedback interaction in which the output concentration regulates the sensor production rate. Positive or negative autoregulations may be present in the sensor and output components. The strength of the interaction between two components ij is quantified by its pairwise susceptibility s_{ij}.

The module described in Fig.(3.1) is modeled by the following system of ordinary differential equations (ODE)

$$\begin{aligned}\frac{dn_I}{dt} &= \alpha_I[1 + a_I \sin(\omega_I t)] - \delta_I n_I \\ \frac{dn_S}{dt} &= \alpha_S F_S(n_I, n_S, n_O) - \delta_S n_S \\ \frac{dn_O}{dt} &= \alpha_O F_O(n_I, n_S, n_O) - \delta_O n_O \end{aligned} \tag{3.1}$$

where n_i states for the number of molecules of each species. In order to study propagation of oscillatory signals, the input may oscillate with frequency ω_I if $a_I \neq 0$. α_i are the maximal production rates for each component (e.g., promoter strengths if the components are transcription factors), and degradation is assumed linear with rates $\delta_i = \tau_i^{-1}$. The production terms F_S and F_O are given in terms of Hill functions, f_{act} and f_{rep} for activation and repression respectively

$$f_{act}(n_i, k_i) = \frac{(n_i/k_i)^H}{1 + (n_i/k_i)^H} \tag{3.2}$$

$$f_{rep}(n_i, k_i) = \frac{1}{1 + (n_i/k_i)^H} \tag{3.3}$$

where k_i is the corresponding activation/repression threshold for ith species and H the Hill coefficient. We already introduced these type of function in the first chapter Eq. (1.8), where we showed that protein production rate follows a sigmoidal curve depending on transcription factor concentration Fig.(1.8). Hill functions are used here to modelize individual interactions between genes. When the expression level of one gene (S or O, depending on the case) is regulated by two different transcription factors, combinations of Hill functions account for the logic of the process: 'OR' regulatory logic is implemented by the sum of the Hill functions assigned to each interaction, while an 'AND' gate is modeled by their product. The behavior of two activating interactions for both 'AND' and 'OR' gates is illustrated in Fig.(3.2).

Depending on the expression of F_S and F_O, we can modelize a broad set of circuit architectures. Some of the most relevant ones are presented in Fig.(3.3), and the form of F_S and F_O is given in Table (3.1).

3.2.2. Quantifying interaction strengths: susceptibilities

To quantify the strength of the interactions between the different species in the network, we take advantage of the concepts of elasticity and susceptibility. Given a network of N species interacting with each other, the time evolution of each species number of molecules n_i can be modeled by

$$\frac{dn_i}{dt} = J_i^+(n_1, n_2, ..., n_N) - J_i^-(n_1, n_2, ..., n_N) \tag{3.4}$$

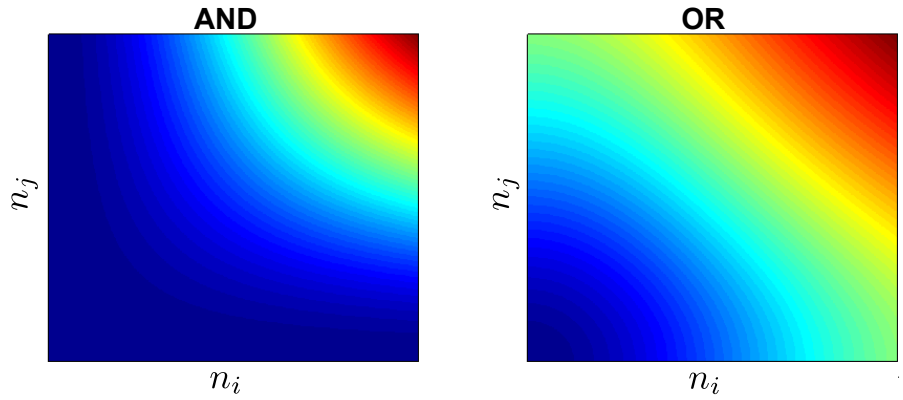


Figure 3.2.: Gene interaction with two different logic gates: both i and j species activate a target gene. For an 'AND' gate interaction between i and j, expression is enhanced (red) when the concentration $n_{i(j)}$ of both species is high, while it is kept at low levels (blue) when either n_i or n_j are low. On the contrary, for an 'OR' gate, expression increases as soon as the concentration of i or j increases, reaching its maximum value when both activating species work together.

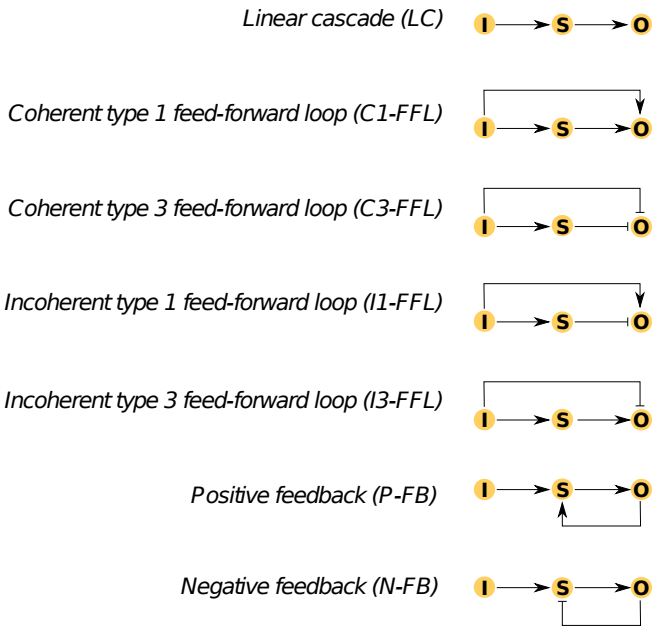


Figure 3.3.: Circuit architectures.

where J_i^+ and J_i^- are defined respectively as the production and degradation rates or fluxes of the ith species. If we consider Eq.(3.4) at steady state, we have

$$\bar{J}_i^+ = \bar{J}_i^- = \bar{J} \tag{3.5}$$

where upper bars denote steady state averages.

Following the definitions provided in previous works [34, 108], we define the pairwise

	AND		OR	
	F_S	F_O	F_S	F_O
LC	$f_{act}(n_I)$	$f_{act}(n_S)$	$f_{act}(n_I)$	$f_{act}(n_S)$
C1-FFL	$f_{act}(n_I)$	$f_{act}(n_S) \cdot f_{act}(n_I)$	$f_{act}(n_I)$	$f_{act}(n_S) + f_{act}(n_I)$
C3-FFL	$f_{act}(n_I)$	$f_{rep}(n_S) \cdot f_{rep}(n_I)$	$f_{act}(n_I)$	$f_{rep}(n_S) + f_{rep}(n_I)$
I1-FFL	$f_{act}(n_I)$	$f_{rep}(n_S) \cdot f_{act}(n_I)$	$f_{act}(n_I)$	$f_{rep}(n_S) + f_{act}(n_I)$
I3-FFL	$f_{act}(n_I)$	$f_{act}(n_S) \cdot f_{rep}(n_I)$	$f_{act}(n_I)$	$f_{act}(n_S) + f_{rep}(n_I)$
P-FB	$f_{act}(n_I) \cdot f_{act}(n_O)$	$f_{act}(n_I)$	$f_{act}(n_I) + f_{act}(n_O)$	$f_{act}(n_I)$
N-FB	$f_{act}(n_I) \cdot f_{rep}(n_O)$	$f_{act}(n_I)$	$f_{act}(n_I) + f_{rep}(n_O)$	$f_{act}(n_I)$

Table 3.1.: Regulation functions for different logic gates: for simplicity, threshold constants k_i are not included in the descriptions, $f_{act(rep)}(n_i) \Leftrightarrow f_{act(rep)}(n_i, k_i)$. Functions for autoregulation interactions in one species can be easily implemented by adding the activating or repressing Hill function for the same species.

elasticity H_{ij} between two species ij as

$$H_{ij} = -\frac{\bar{n}_j}{\bar{J}_i}\left(\frac{\partial \bar{J}_i^+}{\partial \bar{n}_j} - \frac{\partial \bar{J}_i^-}{\partial \bar{n}_j}\right) \tag{3.6}$$

Similarly, we define the pairwise susceptibilities of the interactions as the relative change in species i at equilibrium after a change in species j occurs

$$s_{ij} = -\frac{\bar{n}_j}{\bar{n}_i}\frac{d\bar{n}_i}{d\bar{n}_j} \tag{3.7}$$

As demonstrated in [108], a simple relation can be found between elasticities and susceptibilities

$$s_{ij} = -\frac{H_{ij}}{H_{ii}} \tag{3.8}$$

Elasticities and susceptibilities can be both interpreted in a similar way: they account for the strength of the interactions between species in the network, in the sense that they both quantify the effect that a change in the number of species j has on the equilibrium value of species i. In absolute value, the higher H_{ij} and s_{ij}, the stronger the interaction. Accordingly, their sign depends on the effect of that change: if species i and j increase together, then $H_{ij} < 0$ and $s_{ij} > 0$, while signs change if the species have opposite tendencies. The terms H_{ii} reflect the effect of autoregulatory interactions. If there is no autoregulation of species i, $H_{ii} = -1$ and $s_{ii} = 1$. For positive autoregulation $-1 < H_{ii} < 0$, while $H_{ii} < -1$ if that interaction is negative. For autoregulations the picture changes a little: if no autoregulation is present, then $H_{ii} = s_{ii} = 1$, with this quantity increasing or decreasing depending on the sign and strength of the autoregulation interaction. Thus, we note that, when no autoregulations are found, the relationship between elasticities and susceptibilities becomes just $s_{ij} = -H_{ij}$.

In general, in the following sections we will always talk about susceptibilities when referring to the strengths of the interactions, although some theoretical expressions will be expressed in terms of elasticities. This is due to the fact that, on the one hand, the sign of the susceptibilities matches the sign of the interaction they account for, allowing for an easier understanding and discussion on the observed phenomena. On the other hand, the Jacobian matrix of the system can be easily expressed in terms of elasticities [34, 108], leading to theoretical approximations being written in terms of those quantities rather than susceptibilities. Nevertheless, we have shown that the both quantities are closely related, thus no confusion should be expected.

3.2.3. Quantifying signal detection and propagation

The characterization of the responses that simple gene modules of different topology present to diverse signals is a key step in the understanding of the function of global gene networks. In addition, it allows to develop reverse engineering techniques, which help to unveil the topology of a circuit without interfering with its working machinery, just by comparing the circuit response to complex stimuli [109–112]. For example, Mettetal and co-workers used the FM response of osmo-adaptation in yeast to infer the existence of a negative feedback loop in the system. Moreover, understanding how molecular circuits deal with noisy signals is of key importance in the global understanding of robustness and adaptation in biological systems. In addition, noise itself has proven to be useful to again characterize the structure of gene circuits [113].

Inspired by this ideas, we characterize the response of diverse simple three component modules Fig.(3.1) to AM and FM signals. Then, we focus on the response of these modules to noisy signals.

A framework to study noiseless AM and FM signal detection

Before introducing noise in our analysis, we investigate how AM and FM signals are detected and propagated along circuits of various topologies. Given the simple module described in Fig.(3.1), AM signals are easily modeled as an increase in the input concentration. This leads to a change in output concentration whose amplitude depends on the interactions of the circuit and their strengths. On the other hand, FM detection is studied by stimulating the circuit with an oscillatory input. Given the ODE system in Eq.(3.1), when $a_I > 0$, the input oscillates with frequency ω_I and amplitude proportional to a_I. The oscillation propagates along the network, causing the output to oscillate around its equilibrium value n_O with frequency ω_I, a certain phase-lag, and an amplitude $A(\omega_I)$ which depends on ω_I and, again, the topology of the network Fig.(3.4).

Two different quantities are used to measure the AM and FM detection capabilities of the network. To quantify the potential of the module to detect amplitude variation, we used the

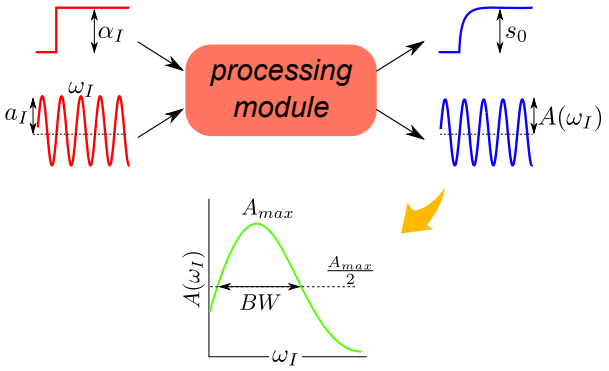


Figure 3.4.: AM and FM signal transmission: depending on the circuitry inside the processing module, AM and FM signals are transmitted with more or less accuracy. AM propagation is quantified by the total susceptibility s_O, while FM transmission is characterized by the bandwidth of the propagated oscillations BW.

output susceptibility s_O [34, 108, 114]. As we already described in the *Methods* section, this quantity measures the relative change in the number of output molecules, n_O, as the input signal changes. The larger s_O the larger the change in the output and therefore, the better the detection. s_O can be written in terms of the pairwise susceptibilities between module components s_{ij} [108, 115]

$$s_O = \frac{n_I}{n_O}\frac{dn_O}{dn_I} = \frac{s_{OI} + s_{OS}s_{SI}}{1 - s_{OS}s_{SO}}. \tag{3.9}$$

On the other hand, to estimate FM detection abilities, we focus on the amplitude of the oscillations propagated along the network, as given by Eq.(3.35). The relative amplitude $A(\omega_I)$ is directly used to compute the bandwidth of the propagated oscillations, which provides a good estimate of the range of frequencies which are properly transmitted along the network . $A(\omega_I)$ is close to its maximum value for a given frequency range and then decreases Fig.(3.4). Its bandwidth is therefore defined as the range of input frequencies where $A^2(\omega_I) > A^2_{max}/2$. The larger this score, the more frequencies are transmitted. Although Eq.(3.35) is strictly valid in the limit of small signal amplitudes, it reproduces quite accurately the response to signals changing up to 50% the equilibrium values Fig.(3.7).

A framework to study noisy AM and FM signal detection

Studying the ability of gene networks to propagate noiseless signals gives us a first insight on the transmission properties associated with the different interactions (feedbacks, feedforwards). However, the picture is not complete unless noise is properly taken into account. Fluctuations of the signal and the rest of the module components may easily disguise the information encoded in both AM and FM signals Fig.(3.5). Therefore, we need new quantities to determine the ability of the different modules to propagate information while dealing

with the unavoidable fluctuations. Thus, we define the signal-to-noise ratio (SNR) for both AM and FM signal propagation. For AM detection, we express the SNR as

$$SNR_{amp} = \frac{s_O}{\sigma_O}, \tag{3.10}$$

where s_O is again the output susceptibility and σ_O the noise coefficient of variation that we have already defined in the *Methods* section Eq.(3.21). SNR_{amp} reflects how detection of signal amplitude may be corrupted by the relative fluctuation amplitude in the output Fig.(3.5).

Defining a SNR for FM signals is less intuitive. Although oscillations could in principle be masked by large amplitude fluctuations [116], we can correctly assume that, since in this case information is encoded in the frequency domain, noise will only affect if its amplitude at the transmission frequency is too high. Given a noisy oscillatory signal, its power spectrum consists of a peak at the oscillatory frequency ω_I and a noisy background Fig.(3.5). Therefore, we can define a SNR for FM signals as the ratio between the amplitude of this peak and the amplitude of the background at the input frequency. This is equivalent to calculating the ratio between the relative amplitude of the propagated oscillations $A^2(\omega_I)$, given by Eq.(3.35), and the amplitude of the fluctuations, given by the power spectrum $P_{fluc}(\omega_I)$ Eq.(3.28) Fig.(3.5).

$$SNR_{freq}(\omega_I) = \frac{A^2(\omega_I)}{\pi P_{fluc}(\omega_I)}. \tag{3.11}$$

3.2.4. Linear approximations

To characterize the response of our model Fig.(3.1) and understand how the different interactions affect its behavior when stimulated with inputs of diverse nature, we can either numerically solve the ODE system presented in Eq.(3.1) (if we are interested in deterministic results), or perform Gillespie simulations of the system [45] to take into account fluctuations due to low copy numbers and reaction probabilities. Although these approaches are in principle enough to characterize the system, they lack for a detailed and systematic description of the influence that the different circuit features have on the observed responses. Thus, to better understand the behavior of this simple three component module, we develop theoretical predictions for the different responses analyzed during our study, based on linear approximations around the steady state. These approximations will allow us to relate specific responses to the network structure, quantified by the strength of its interactions, and to a few key features of the network, such as degradation rates.

To start, we define the relative deviation from equilibrium ($\bar{n}_i$) of a given species number as

$$\Delta n_i = \frac{n_i - \bar{n}_i}{\bar{n}_i} \tag{3.12}$$

Then, we can take the system in Eq.(3.4), perform a first order linearization [117], and rewrite it in terms of these relative deviations from equilibrium. By doing so we end up with a linearized ODE system for the time evolution of the deviations

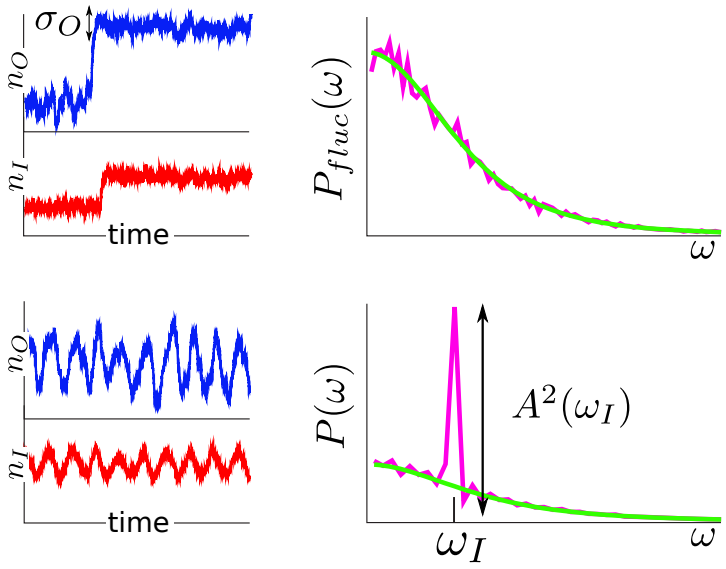


Figure 3.5.: AM and FM noise properties: a noisy input signal (red) propagates through the network causing a response in the output (blue). The output noise amplitude is characterized by its coefficient of variation σ_O and its frequency content $P_{fluc}(\omega)$, Eq.(3.28), -simulation result (pink), and theoretical prediction (green)-. Given a noisy oscillatory input of frequency ω_I, the output responds with an oscillation whose power spectra $P(\omega)$ has a background corresponding to $P_{fluc}(\omega)$ and a narrow peak of height $A^2(\omega_I)$, Eq.(3.35), at the input frequency.

$$\frac{d\Delta \boldsymbol{n}}{dt} = \boldsymbol{M} \cdot \Delta \boldsymbol{n} \tag{3.13}$$

Here, $\Delta \boldsymbol{n}$ is the column vector containing all the Δn_i elements corresponding to the different species. The matrix $\boldsymbol{M}$ is the Jacobian matrix of the system, whose elements are given by

$$M_{ij} = \frac{\bar{n}_j}{\bar{n}_i}\left(\frac{\partial \bar{J}_i^+}{\partial \bar{n}_j} - \frac{\partial \bar{J}_i^-}{\partial \bar{n}_j}\right) \tag{3.14}$$

Comparing this expression with the definition of pairwise elasticities Eq.(3.6), we can write the M_{ij} elements like

$$M_{ij} \equiv -\frac{H_{ij}}{\tau_i}. \tag{3.15}$$

where $\tau_i \approx \bar{n}_i / \bar{J}_i$ states for turnover rate of the *ith* species.

Applying all these definitions to our particular three species module Eq.(3.1), we can write the Jacobian matrix $\boldsymbol{M}$ as

$$\boldsymbol{M} = -\begin{pmatrix} \frac{H_{II}}{\tau_I} & 0 & 0 \\ \frac{H_{SI}}{\tau_S} & \frac{H_{SS}}{\tau_S} & \frac{H_{SO}}{\tau_S} \\ \frac{H_{OI}}{\tau_O} & \frac{H_{OS}}{\tau_O} & \frac{H_{OO}}{\tau_O} \end{pmatrix} \tag{3.16}$$

where the characteristic times are now given by $\tau_i = \delta_i^{-1}$, being δ_i the degradation rates in Eq.(3.1).

The stability of the fixed points of the system is given by the sign of the determinant of the Jacobian matrix

$$Jac \equiv \frac{1}{\tau_S \tau_O}(H_{SS}H_{OO} - H_{SO}H_{OS}) = \frac{H_{SS}}{\tau_S}\frac{H_{OO}}{\tau_O}(1 - s_{SO}s_{OS}) \tag{3.17}$$

and its trace

$$Tr \equiv -\left(\frac{H_{SS}}{\tau_S} + \frac{H_{OO}}{\tau_O}\right). \tag{3.18}$$

With this linearization around the steady state of the original three component system, we can derive approximate expressions for the responses we want to characterize in our study, namely the amplitude of the system fluctuations and their power spectrum, the amplitude of oscillations propagated along the circuit, and the dynamics of small deviations from the equilibrium state. All these expressions are ultimately written in terms of susceptibilities or elasticities and degradation rates, allowing for a deep analysis of the effect of each interaction strength in the overall response of the system.

Noise amplitude

The reinterpretation of the Fluctuation-Dissipation Theorem (FDT) made by J. Paulsson [114, 118] allows us to obtain solutions for the time evolution of the covariance matrix of the fluctuations $\boldsymbol{\sigma}$ in networks of interacting molecules. At equilibrium, the covariance can be obtained by solving the following Lyapunov equation

$$\boldsymbol{M} \cdot \boldsymbol{\sigma} + \boldsymbol{\sigma} \cdot \boldsymbol{M}^{\boldsymbol{T}} + \boldsymbol{D} = \boldsymbol{0} \tag{3.19}$$

Here, $\boldsymbol{M}$ states for the Jacobian matrix of the system Eq.(3.15). $\boldsymbol{D}$ is the diffusion matrix, and is composed of the noise sources of the system. It depends on the reaction fluxes, the system size and the stoichiometry of the reactions. Given that each reaction affects only one molecular species ($D_{ij} = 0$, $\forall j \neq i$), and that a single molecule is added or removed in each reaction, the elements of the diffusion matrix take the form

$$D_{ii} = \frac{2}{V_i \bar{n}_i}\frac{1}{\tau_i} \tag{3.20}$$

being V_i the *ith* species effective volume, which in our specific case will be used to control each species intrinsic noise. This will allow us to test how intrinsic noise from different components in the network contributes to the overall fluctuations of the system.

Solving Eq.(3.19) for our specific three species model, we obtain an expression for each species coefficient of variation (noise amplitude) by looking at the diagonal terms of the matrix $\boldsymbol{\sigma}$. In this case, the output noise amplitude is comprised of the sum of three terms

σ^2_{OK}, each one representing the propagation along the circuit of the intrinsic fluctuations of each species $K = I, S, O$

$$\sigma^2_O = \sigma^2_{OO} + \sigma^2_{OS} + \sigma^2_{OI} \tag{3.21}$$

The terms in Eq.(3.21) are given by

$$\begin{aligned} \sigma_{OO} &= \sigma^{int}_O \left(1 + s_{SO} \frac{H_{OS}/\tau_O}{T_{av}}\right) \\ \sigma_{OS} &= \sigma^{int}_S s^2_{OS} \frac{H_{OO}/\tau_O}{T_{av}} \\ \sigma_{OI} &= \sigma^{int}_I \left(s^2_O \frac{T^{tot}_{av}}{T_{av}} \frac{Jac}{Jac + \frac{H_{II}}{\tau_I} T^{tot}_{av}} + s^2_{OI} \frac{H_{OO}/\tau_O}{T_{av}} \frac{\frac{H_{II}}{\tau_I}\frac{H_{OO}}{\tau_O}}{Jac + \frac{H_{II}}{\tau_I} T^{tot}_{av}}\right) \end{aligned} \tag{3.22}$$

being

$$\begin{aligned} T^{tot}_{av} &= H_{II}/\tau_I + H_{SS}/\tau_S + H_{OO}/\tau_O \\ T_{av} &= -Tr = H_{SS}/\tau_S + H_{OO}/\tau_O \end{aligned} \tag{3.23}$$

Each term in Eqs.(3.21) is proportional to each species intrinsic noise σ^{int}_K

$$\begin{aligned} \sigma^{int}_I &= \frac{D_{II}\tau_I}{2} \\ \sigma^{int}_S &= \frac{D_{SS}\tau_S}{2} \frac{1}{H_{SS} + s_{IS}H_{SI}} \\ \sigma^{int}_O &= \frac{D_{OO}\tau_O}{2} \frac{1}{H_{OO} + s_{SO}H_{OS}} \end{aligned} \tag{3.24}$$

This magnitude is defined as the contribution to the fluctuations of each species produced by its low number of molecules. Thus, it takes the form of the coefficient of variation for a typical Poisson process ($\sigma^{int}_i \sim \frac{D_{ii}\tau_i}{2} \sim \frac{1}{\bar{n}_i}$)

Noise frequency content

Fluctuations in the number of molecules of the different network species are characterized not only by their amplitude (coefficient of variation), but also by their frequency content. This property is given by the power spectra of the fluctuations $P_{fluc}(\omega)$, which quantifies the different frequencies at which the system fluctuates and their contributions to the dynamics of noise. Similarly to the noise coefficient of variation, theoretical expressions for $P_{fluc}(\omega)$ can be written in terms of elasticities and degradation times. To get to this expression, we start writing a Langevin equation which describes the dynamics of our already linearized system and takes into account random fluctuations by adding an extra term ζ_i

$$\frac{d\eta_i}{dt} = \sum_{j=1}^{N} M_{ij}\eta_j + \zeta_i \tag{3.25}$$

The fluctuating term consists of white noise obeying the FDT

$$< \zeta_i(t)\zeta_j(0) >= D_{ij}(\boldsymbol{n})\delta(t) \tag{3.26}$$

where $\boldsymbol{D}$ is, again, the diffusion matrix described in Eq.(3.20).

The power spectrum matrix for the fluctuations can be obtained by solving the following equation [74]

$$P_{\boldsymbol{\eta}}(\omega) = \frac{1}{2\pi}\left(\boldsymbol{M} + \boldsymbol{I}i\omega\right)^{-1}\boldsymbol{D}\left(\boldsymbol{M}^T - \boldsymbol{I}i\omega\right)^{-1} \tag{3.27}$$

where $\boldsymbol{I}$ is the identity matrix.

Thus, by doing some algebra we can solve Eq.(3.27) and obtain an expression for the fluctuations power spectra of the output species in our linearized three component system

$$\begin{aligned} P_{fluc}(\omega) &= D_{OO}\frac{\omega^2 + \frac{H_{SS}^2}{\tau_S^2}}{\Delta(\omega)} + D_{SS}\frac{s_{OS}^2 H_{OO}^2/\tau_O^2}{\Delta(\omega)} \\ &+ \frac{D_{II}}{\left(\omega^2 + \frac{H_{II}^2}{\tau_I^2}\right)}\frac{s_2^2 Jac^2 + s_{20}^2\frac{H_{OO}^2}{\tau_O^2}\omega_I^2}{\Delta(\omega)}. \end{aligned} \tag{3.28}$$

where $\Delta(\omega)$ is given by

$$\begin{aligned} \Delta(\omega_I) &\equiv \left(\omega_I^2 + \frac{H_{SS}^2}{\tau_S^2}\right)\left(\omega_I^2 + \frac{H_{OO}^2}{\tau_O^2}\right) \\ &+ 2s_{SO}s_{OS}\frac{H_{SS}}{\tau_S}\frac{H_{OO}}{\tau_O}\left(\omega_I^2 + \alpha\right), \end{aligned} \tag{3.29}$$

and $\alpha \equiv \left(\frac{s_{SO}s_{OS}}{2} - 1\right)\frac{H_{SS}}{\tau_S}\frac{H_{OO}}{\tau_O}$.

Similarly to the coefficient of variation, each term in this sum depends on the noise intensities of the species involved in the reaction. The first two terms represent the contribution to the fluctuation spectrum of the intrinsic noises of output and sensor species, while the third one comprises the contribution to the overall spectra of the input fluctuations propagated along the network.

Propagated oscillations

To estimate the amplitude of oscillations propagated along the network we consider the general system in Eq.(3.13) and include a term $\boldsymbol{q}(t)$ which states for oscillations in the production rate of each species

$$\frac{d\Delta\boldsymbol{n}}{dt} = \boldsymbol{M}\cdot\Delta\boldsymbol{n} + \boldsymbol{q}(t) \tag{3.30}$$

where the elements of the vector $\boldsymbol{q}(t)$ take the general form

$$q_i(t) = \frac{a_i}{\tau_i}\sin(\omega t + \phi_i) \tag{3.31}$$

As shown by Samoilov and coworkers [119], the amplitude of the stationary solution of Eq.(3.30) is given by

$$A^2(\omega)_{t\to\infty} = (i\omega \boldsymbol{I} - \boldsymbol{M})^{-1}\boldsymbol{\gamma}\cdot\boldsymbol{\gamma}^{\dagger}(-i\omega \boldsymbol{I} - \boldsymbol{M}^T)^{-1} \tag{3.32}$$

where each term in the $\boldsymbol{\gamma}$ matrix is given by

$$\gamma_i = a_i e^{i\phi_i} \tag{3.33}$$

In our three species system all the elements of the time-dependent term are zero except for the first one, which takes the form $q_I(t) = \frac{a_I}{\tau_I}\sin(\omega_I t)$. Therefore,

$$\boldsymbol{\gamma}\cdot\boldsymbol{\gamma}^{\dagger} = \begin{pmatrix} \frac{a_I^2}{\tau_I^2} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{3.34}$$

Thus, by solving Eq.(3.32) we get an estimate for the deviation of the output oscillations from the equilibrium

$$\begin{aligned} A(\omega_I) &\equiv \frac{\max\left[n_O(t)\right] - n_O}{n_O} \\ A^2(\omega_I) &= \frac{a_I^2}{\tau_I^2\left(\omega_I^2 + \frac{H_{II}^2}{\tau_I^2}\right)} \frac{s_O^2 Jac^2 + s_{OI}^2 \frac{H_{OO^2}}{\tau_O^2}\omega_I^2}{\Delta(\omega_I)}. \end{aligned} \tag{3.35}$$

It is important to notice that, as expected, the dependence of the propagated oscillations on the elasticities, susceptibilities and decay rates, matches exactly the third term in Eq.(3.28), which describes the contribution of the input fluctuations to the overall power spectra.

Response-times

So far, all the signal propagation features for which theoretical expressions were derived have been evaluated at equilibrium. Nevertheless, as it will be shown in the next sections, these signal propagation features are strongly linked with the dynamical behavior of the networks. This connection can be investigated computationally, but, as in the previous cases, developing theoretical expressions for the dynamics of the circuits provides with a powerful tool to discriminate the origin of the observed phenomena.

The dynamic response of a network can be approximated by finding analytical solutions for the time evolution of the linearized system Eq.(3.13). No condensed solution can be found for the dynamics of any three component network like the one presented in Fig.(3.1), although it is possible to find it for any two component module consisting only in a sensor and output component Eqs.(B.4,B.7). Nevertheless, Eq.(3.13) can be individually solved for specific three component cases: linear cascades, feedbacks and feedforwards, leading to

the expressions presented in the *appendix* section Eqs.(B.10-B.18). As in the signal propagation cases, the solutions are written in terms of elasticities, susceptibilities and decay rates, building up a common theoretical framework to analyze responses and dynamics of simple modules.

Once analytical approximations are found for the dynamics of different networks, we quantify the output dynamical response by means of the response-times. They can be computed by solving the ODE system or by means of the linear approximations for the response mentioned above. In both cases, we initially consider the system at equilibrium in what we called its off state. At certain point, the input concentration is increased to a new value, triggering a response in the circuit, which evolves to what we called the on state. Then, the input is switched off again, and the system comes back again to the off state. For both transitions between on and off states we can calculate the response times, T_{on} and T_{off}, defined as the time spent by the output to travel half the way to its new equilibrium state upon a change in the input concentration Fig(3.6). During this process, the system always sees the final input amplitude, therefore we linearize it around its on state to compute T_{on}, and around its off state to compute T_{off}

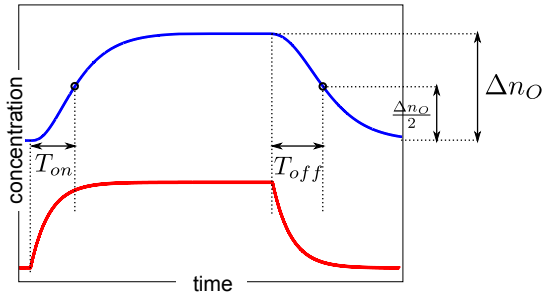


Figure 3.6.: Definition of response-times: upon an increase in the input (red) production rate, the activation time T_{on} is defined as the time spent by the output (blue) to travel half its way to the new equilibrium value. Accordingly, the deactivation time T_{off} is defined as the time spent by the output to travel half the way back to the original concentration when the input production rate is restored.

Given the time-evolution of the output deviations from equilibrium $\Delta n_O(t)$, we can write in terms of the response times that

$$\begin{aligned} \Delta n_O^{on}(T_{on}) &= \frac{\Delta n_O^{on}(0)}{2} \\ \Delta n_O^{off}(T_{off}) &= \frac{\Delta n_O^{off}(0)}{2} \end{aligned} \tag{3.36}$$

where, as usual

$$\Delta n_O^{on/off}(t) = \frac{n_O^{on/off}(t) - \bar{n}_O^{on/off}}{\bar{n}_O^{on/off}} \tag{3.37}$$

being $\bar{n}_O^{on/off}$ the equilibrium value of the output in its on or off state.

The fact that T_{on} and T_{off} are computed linearizing around different equilibrium states implies that, unlike for the steady state responses of the output to different input signals, the dynamic response of the system is governed not only by the susceptibilities and decay rates, but also by the specific off and on concentrations of the different species. As it will be shown in the next sections, this causes circuits with the same topology but different cis-regulatory functions to exhibit rather different dynamical properties.

Testing linear approximations

To check the validity of the previous linear approximations we compared analytical predictions with results obtained from simulations. Responses to an oscillatory input are computed by solving the ODE system in Eq.(3.1) with a standard Runge-Kutta algorithm. Noise amplitude and power spectra are obtained from Gillespie simulations, while response times are obtained by characterizing the evolution of the ODE system upon changes in the input concentration.

Fig.(3.7) shows the results for specific parameter values, although similar results are obtained for a broad set of different susceptibilities and species concentrations and volumes. We see that both analytical solutions and simulations are in very good agreement, giving us confidence in the results obtained from the linear approximations.

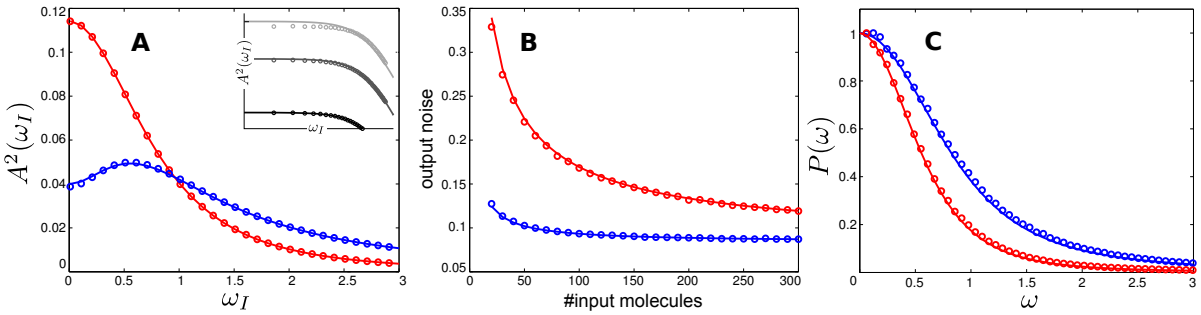


Figure 3.7.: Testing linear approximations for the amplitude of propagated oscillations (A), noise coefficient of variation (B) and fluctuations power spectra (C). Results for both C-FFL (red) and I-FFL (blue) are presented. Theoretical approximations (lines) agree with simulation results (points). The inset in A shows the validity of the approximations when the input oscillations increase in amplitude: $a_I = 0.01$ (black), $a_I = 0.1$ (dark grey) and $a_I = 0.5$ (light grey). The rest of the parameters are: number of input molecules 200, sensor and output 400. Susceptibilities: $s_{SI} = 1$, $s_{OS} = 2$, $s_{OI} = \pm 1.5$.

The results of testing the analytical predictions for the dynamics of the circuits are presented in Fig.(3.8) for a two species LC. Response times obtained from simulations and from theoretical predictions are compared as a function of the threshold constant K. Although the accuracy of the theoretical predictions is not perfect, linear approximations are still able to reproduce the overall tendency of the system, including regions of positive and negative

slope. Similar results are obtained for other types of circuits and logic gates.

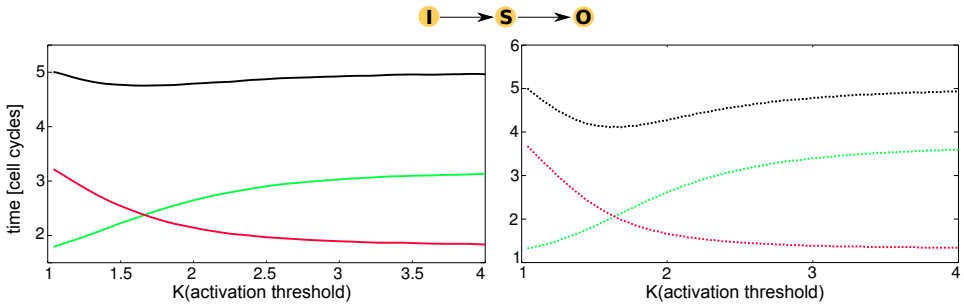


Figure 3.8.: Testing linear approximations for the response times of a LC. Theoretical predictions (dashed lines) are compared with simulation results (solid lines). T_{on} is represented in green, T_{off} in red, and ΣT in black.

3.2.5. Statistical analysis

In the second part of the *Results* section bellow, we will show the influence that the structure of the circuit has on the ability of the circuit to propagate certain signals, as well as the constraints it imposes in the dynamic response. This study is based on a statistical analysis which allows to extract general conclusions which do not depend on the specific regime the circuits are operating in (susceptibilities and decay rates). Three different tools are used to compute correlations between data sets and extract useful information: Spearman's rank correlation (SC), Mutual information coefficient (MIC) and principal component analysis (PCA).

Spearman's rank correlation

SC describes how well the relationship between two variables can be described by a monotonic function. Unlike the usual Pearson's correlation coefficient, SC is able to deal with non-linear relationships between variables, being therefore more flexible when analyzing data sets whose relationship is, in principle, unknown. Given N pairs of raw scores (X_i, Y_i), to calculate SC we first compute the ranked variables (x_i, y_i), defined as the position (in descending order) that each raw score (X_i or Y_i) occupies in its corresponding set (ranks for X's and Y's are computed independently). Then, $SC(X, Y)$ is computed as the standard Pearson's correlation coefficient of the ranked data (x_i, y_i)

$$SC(X,Y) = \frac{\Sigma_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\Sigma_i (x_i - \bar{x})^2 \Sigma_i (y_i - \bar{y})^2}} \tag{3.38}$$

where $\bar{x}$ and $\bar{y}$ respectively state for x's and y's averages. SC takes values between -1 and 1, with positive and negative values respectively indicating that Y increases or diminishes with X. The larger SC (in absolute value), the stronger the correlation.

Mutual information coefficient

The second quantity we use to quantify correlations between variables is the Mutual information coefficient. This quantity has been previously applied to reveal form-function constraints in biological networks [120]. In contrast with SC, it has the advantage that is able to unveil non-monotonic relations between variables, although it does not distinguish whether Y increases or diminishes with X. Thus, it takes values between 0 and 1 as the correlation increases.

The Mutual information $I(X,Y)$ between two variables is defined as

$$I(X,Y) = \sum_X \sum_Y p(X,Y) log\left(\frac{p(X,Y)}{p(X)p(Y)}\right) \tag{3.39}$$

being $p(i)$ and $p(X,Y)$ the marginal and joint probability density functions respectively.

From there, we can normalize $I(X,Y)$ to define the Mutual information coefficient $MIC(X,Y)$

$$MIC(X,Y) = \frac{I(x,y)}{min(H(x),H(y))} \tag{3.40}$$

where $H(i)$ is the entropy of feature i, given by

$$H(i) = -\sum_i p(i) log(p(i)) \tag{3.41}$$

Although in principle other normalization criteria could be used [121], we decide to choose the one described in Eq.(3.40) because it is symmetric, $MIC(X,Y) = MIC(YX)$, and has been previously used in similar analysis, like the one performed by Mugler we cited above.

Principal component analysis

Unlike SC or MIC, which account for correlations between pairs of variables, PCA allows us to find correlations between multiple observations. Given a set of N observations $O_1, O_2, ..., O_N$, we build an adjacency matrix $\boldsymbol{A}$ whose elements A_{ij} are the correlations between pairs of observations: $A_{12} = SC(O_1, O_2)$ or $A_{12} = MIC(O_1, O_2)$ Fig.(3.9-A). Then, we can diagonalize $\boldsymbol{A}$ to obtain an orthogonal projection of the previously correlated data in which the new variables, known as principal components, are completely uncorrelated.

The first principal component is defined as the one with higher eigenvalue, and accounts for the largest variance between the original data. The rest of the principal components are defined accordingly to this criteria, in descending eigenvalue order. The associated N-dimensional eigenvectors contain the coordinates in the principal component space of each of the N observables. Thus, if we take, for instance, the first 2 principal components, we can represent a 2D projection of the positions of the observations in the N-dimensional principal component space Fig.(3.9-B). This projection is done over the plane defined by the coordinates which better explain the variability of the system, and allows to visualize correlations between observations: the closer the observations the more correlated they are.

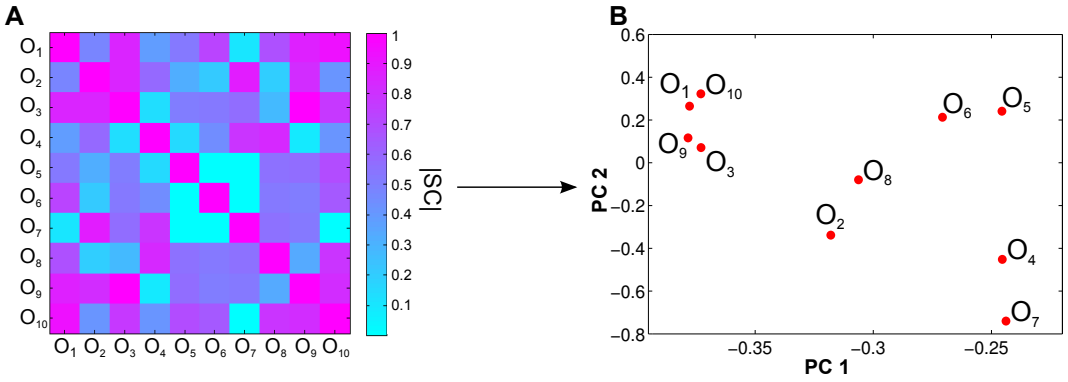


Figure 3.9.: Principal component analysis: correlations (here, absolute value of SC, $|SC|$) between 10 observations O_i are computed to build the adjacency matrix (A). Performing a principal component analysis allows to plot a 2D projection of the observations in the principal components space (B). Observations which lie close in this projection, like O_1 and O_{10}, or O_9 and O_3, are correlated with each other.

3.3. Results

In the previous section we have provided a description of the module used to analyze the signal response properties of simple gene circuits. We also introduced quantities which allow us to quantify the response of the module to both noisy and noiseless AM and FM signals. In addition, we presented a theoretical framework which gives good predictions for the behavior of those systems and which allows for its systematic study. Finally we presented the statistical tools used to analyze the signal response features.

At this point, we are ready to discuss the main results obtained with our analysis. First, we will investigate key signal detection and noise filtering capabilities of relevant gene networks. To finish, we will describe the results of a detailed study of the relationships between the ability of the circuits to propagate diverse signals, their response to fast changes in input concentration and their topology.

3.3.1. Signal detection and noise filtering

The first results we present here focus on the signal detection capabilities of simple gene networks. We show how trade-offs between AM and FM signals emerge for some circuit architectures, while others are able to overcome this limitation. In addition to signal propagation, noise filtering is incorporated in the study, allowing us to find the circuits which exhibit the best noise-filtering capabilities for both AM and FM signals.

Detection of noiseless signals by simple modules

In the *Methods* section we described how AM and FM detection are quantified. Thus, we now investigate how the module structure influences the transmission of both types of signals.

To address this problem we start with the simplest configuration of our three component module: the linear cascade Fig.(3.3). In this case, the output susceptibility is given by $s_O = s_{OS}s_{SI}$, and the amplitude of the output in response to a sinusoidal signal by

$$A^2(\omega_I) = \frac{a_I^2}{\tau_I^2\left(\omega_I^2 + \frac{H_{II}^2}{\tau_I^2}\right)} \frac{s_O^2 Jac^2}{\left(\omega_I^2 + \frac{H_{SS}^2}{\tau_S^2}\right)\left(\omega_I^2 + \frac{H_{OO}^2}{\tau_O^2}\right)}, \tag{3.42}$$

In this case, $A^2(\omega_I)$ is the product of three low-pass filters with individual bandwidths given uniquely by the lifetimes of each component. Since bandwidth does not depend on susceptibilities, frequency detection can be adjusted independently of amplitude detection. Therefore, AM detection can be controlled by tuning s_{OS} and s_{SI}, while FM transmission is determined by τ_i's. It also becomes clear in this framework that the slowest time scale of the system strongly determines its bandwidth [109] Fig.(3.10A), and that adding successive components to the cascade reduces its output bandwidth Fig.(3.10B). This implies that longer cascades filter out more easily transient stimuli, as it has been experimentally shown in a synthetic genetic cascade [122].

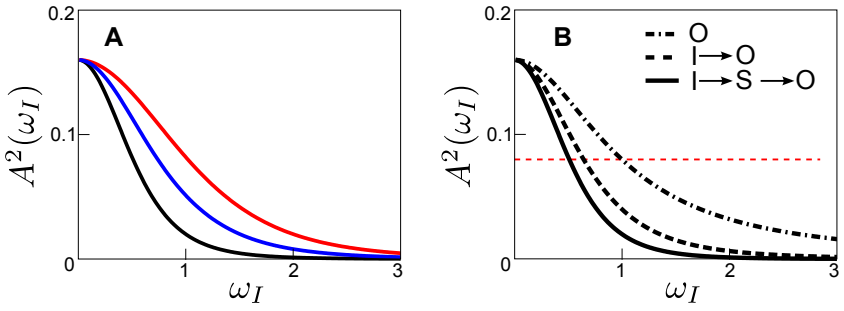


Figure 3.10.: FM detection for various LC's. Dependence of relative amplitude A on species degradations (A): compared to the case in which $\delta_I = \delta_S = \delta_O = 1$ (black), the bandwidth increases when $\delta_S = \delta_O$ increase to 2 (blue) and increases even more if all three species speed up their degradation time $\delta_I = \delta_S = \delta_O = 2$ (red). Bandwidth dependence on LC's length: compared to the case in which a single species oscillates due to oscillations of its production rate (dot-dashed line), bandwidth decreases when more components are added, a second one (dashed line), then a third one (solid line).

What happens if we add to the LC a feedback (FB) interaction from the output to the sensor component? In this case $s_{SO} \neq 0$, and the output susceptibility is given by

$$s_O = \frac{s_{OS}s_{SI}}{1 - s_{OS}s_{SO}}, \tag{3.43}$$

FB interactions can be either positive (P-FB) or negative (N-FB), depending on the sign of the product $s_{OS}s_{SO}$. Now, if we look at Eq. (3.35) we see that in this case $A(\omega_I)$ is similar to LC's except for the third term in the denominator Eq. (3.29). For fixed degradation rates and

no autoregulation of the module components ($H_{SS} = H_{OO} = 1$), this term depends precisely on the product $|s_{OS}s_{SO}|$, which we denote as *feedback strength* (FS).

To determine how FS influences detection, we fix the input-sensor interaction ($s_{SI} = 1$) and the degradation rates ($\tau_I = \tau_S = \tau_O = 1$), while allowing the FB susceptibilities (s_{OS} and s_{SO}) to vary within a range $[smin, smax]$ [108]. This produces a set of FB detection modules with FS in the interval $[smin^2, smax^2]$ (for P-FB, FS<1 for the equilibrium point to be stable).

The results of this analysis are plotted in Fig.(3.11), where we simultaneously plotted bandwidth and output susceptibility as a function of FS for negative and positive FB's. The first result is indeed that there is a one-to-one correspondence between FS and bandwidth, meaning that FS uniquely determines the ability of FB circuits to propagate oscillatory signals. On the other hand, a relatively broad range of susceptibilities can be achieved for a single FS. This is caused by the fact that the same FS can be obtained for different values of the individual susceptibilities s_{OS} and s_{SO}. Now, if we compare the ability of positive and negative FB's to propagate AM and FM signals we notice that they exhibit opposite behaviors. P-FB interactions increase s_O, while, on the contrary, N-FB's diminish that quantity. Thus, P-FB interactions increase the ability of the circuit to detect AM signals, while N-FB's decrease it. In addition, the effect that increasing FS has on the FM detection abilities of both types of circuits is again different: an increase in FS due to a P-FB interaction increases the bandwidth, while it is reduced if that interaction is negative. Therefore, the detection of FM signals is enhanced by negative FB interactions and drops for positive ones. Taking all that into account, we see a clear trade-off in AM and FM detection by FB circuits: although positive and negative FB interactions exhibit opposite trends, they are both constrained by the fact that they can only increase the detection of one type of signal, either AM (P-FB) or FM (N-FB). We have observed a similar tendency in linear cascades with autoregulated components as a function of autoregulation strength.

Having analyzed the effect of adding FB interactions to LC's, we can now study the influence of another interaction which is commonly found in gene circuits: the feed-forward loop interaction (FFL) [66]. It consists of a direct interaction between the input and the output, which again, can be either positive or negative. The output susceptibility in this case is given by

$$s_O = s_{OI} + s_{OS}s_{SI} \tag{3.44}$$

and the amplitude of the oscillatory response follows

$$A^2(\omega_I) = \frac{a_I^2 s_O^2 Jac^2 + a_I^2 s_{OI}^2 \frac{H_{OO^2}}{\tau_O^2}\omega_I^2}{\tau_I^2\left(\omega_I^2 + \frac{H_{II}^2}{\tau_I^2}\right)\left(\omega_I^2 + \frac{H_{SS}^2}{\tau_S^2}\right)\left(\omega_I^2 + \frac{H_{OO}^2}{\tau_O^2}\right)}. \tag{3.45}$$

The first term in $A^2(\omega_I)$ corresponds to a low-pass filter identical to the one found in linear cascades Eq. (3.42). On the other hand, the second term in the numerator, which is proportional to the direct interaction s_{OI}, depends on ω^2, indicating that it describes as a

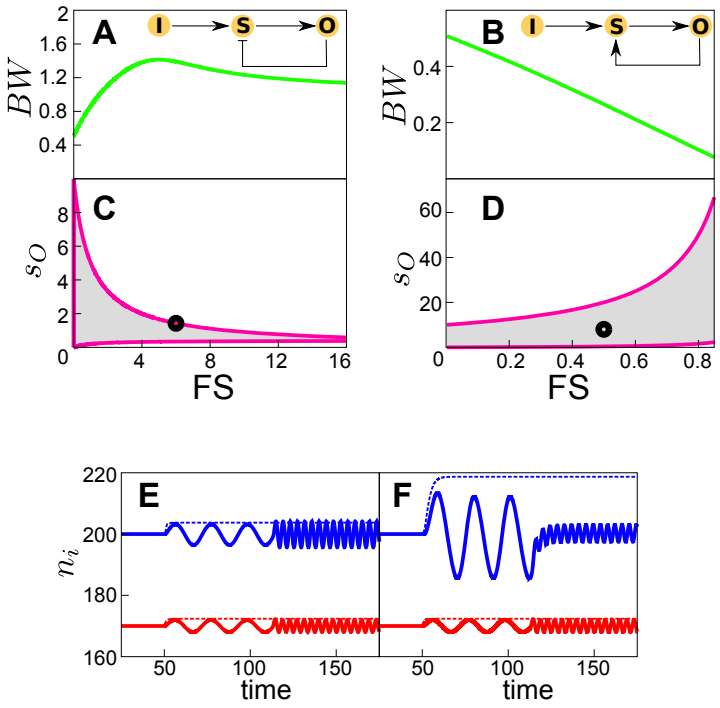


Figure 3.11.: Feedback response to AM and FM signals: the bandwidth of the propagated oscillations BW (green) uniquely depends on FS. A broad range of output susceptibilities s_O can be found for each FS (gray area), although it is limited differently depending on the topology (pink). Boxes E and F respectively show the different responses exhibited by a negative and positive FB working in the region specified by black dots in the above boxes. Respectively in red and blue, the input and output number of molecules n_I and n_O. Dashed lines represent the response to a step like input (AM).

high-pass filter. Therefore, FM detection is given by the competition between low and high-pass filtering.

Similarly to the case of FB circuits, we can find a design attribute which uniquely determines signal detection properties of FFL circuits. We call this quantity *feedforward strength* (FS, which can be easily distinguished from the *feedback strength* by the context), and is given by $FS \equiv |s_{OI}|/s_O|$. To study the influence of adding an extra FFL interaction to the LC's configuration, we follow the same procedure as for FB's, randomly sampling the space of pairwise susceptibilities between $[smin, smax]$ while computing bandwidth and s_O. In Fig.(3.12) we plot the dependence of bandwidth and output susceptibility on FS. We do so for two of the most commonly found FFL motifs: the type 1 coherent FFL (C1-FFL) and the type 3 incoherent FFL (I3-FFL) [66].

In C1-FFL's all the interactions are positive, and therefore $s_O > s_{OI}$. It is clear that increasing s_{OI} increases the overall susceptibility s_O and therefore the AM signal propagation

abilities of the module. This increase in s_O occurs for a significant range of FS's Fig.(3.12-C). Meanwhile, FM detection is always enhanced when FS increases: since the low-pass filtering term depends on s_O, it dominates the overall behavior of the system. However, due to the contribution of the $s_{OI}\omega^2$ high-pass filtering term, FM propagation is always enhanced with respect to the LC configuration Fig.(3.12-A). Therefore, the addition of this positive interaction going directly from the input to the output increases the ability of the circuit to propagate both AM and FM signals.

I3-FFL's exhibit different properties. In this case, similarly to the N-FB, the additional negative interaction reduces the overall circuit susceptibility, thus reducing AM signal detection with respect to LC's Fig.(3.12-D). However, new properties emerge in the FM signal detection side. Although increasing the negative direct interaction (in absolute value) reduces s_O and therefore AM signal propagation, this decrease also reduces the influence of the low-pass filtering term in FM propagation while increasing the high-pass filtering one. In the limit in which ($|s_O| << |s_{OI}|$), the high-pass filter behavior of the module completely dominates FM transmission, leading to a completely new behavior. While the propagation of FM signals in all the previous modules was always better the lower the input frequency of oscillation, in the case of I3-FFL's we can find sets of susceptibilities for which the optimum propagation of the oscillation is found for a frequency greater than 0 Fig.(3.12-B).

To illustrate the behavior of both coherent and incoherent FFL's, the output response to an oscillatory input of increasing frequency is respectively plotted in Figs.(3.12-E and F). As expected, the C1-FFL propagates faster stimuli with less amplitude than slower ones. On the contrary, the I3-FFL working on its high-pass filter regime is able to increase the amplitude of the propagated oscillations when the input frequency increases.

To summarize, we can say that FFL's are flexible AM and FM signal detectors, being able to improve the propagation of both types of signals and work as high-pass filters.

Finally, to generalize the results obtained here for specific a feedforward configuration to any possible three component module, we perform a statistical analysis in which we generate a set of 200,000 modules of random topology and strengths of the interactions. Any possible structure among the 162 configurations consistent with the diagram presented in Fig.(3.1) can be generated, including combinations of feedback, feedforward and autoregulation interactions. For each module, its susceptibility and bandwidth of the propagated oscillations are computed. Based on these results, we select the modules which exhibit better AM and FM detection properties than the linear cascade i.e. higher susceptibility and bandwidth. To finish, we classify the resulting modules based on their topology, and give them a *recurrence* score. This score is the proportion of selected modules of a given type found in our sample compared with the proportion expected if the sample would be random. The result of this process is that, surprisingly, the only motif which is significantly able to show both better AM and FM signal detection properties than LC's is a pure coherent FFL -here, any type of coherent FFL structure is taken into account [63]-, with no additional interactions present

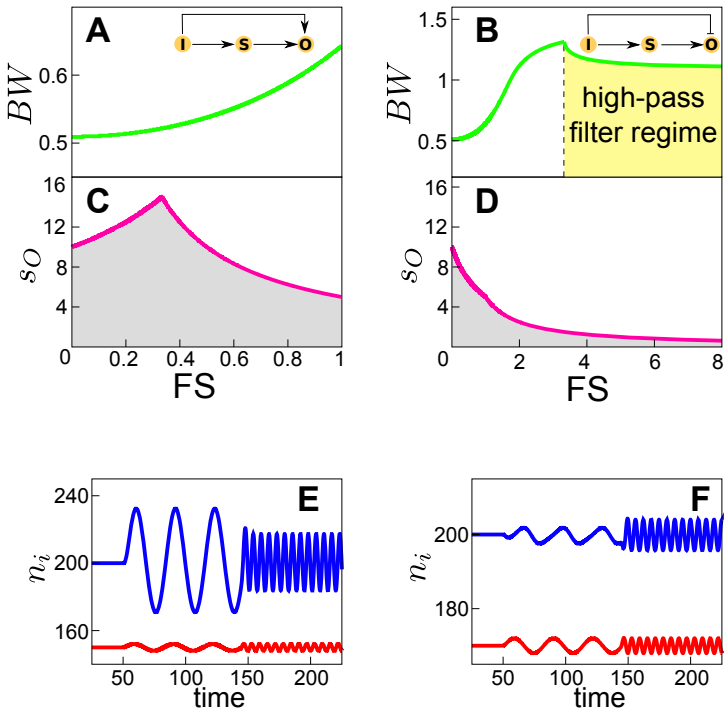


Figure 3.12.: Feedforward responses to AM and FM signals: the bandwidth of the propagated oscillations BW (green) uniquely depends on FS. A broad range of output susceptibilities s_O can be found for each FS (gray area), although it is limited differently depending on the topology (pink). Boxes E and F respectively show the different responses exhibited by a coherent and incoherent FFL. Respectively in red and blue, the input and output number of molecules n_I and n_O. Notice how the I-FFL (F) is capable of showing oscillations of larger amplitude for a faster stimulus.

Fig.(3.13).

Noise-tolerant signal detection

Following the same steps as in the study of the response to noiseless signals, in the *Methods* section we have defined SNR's as the main features to characterize the propagation of noisy signals. Now, we investigate the influence of feedback and feedforward interactions on the transmission of signals perturbed by random fluctuations in comparison to a simple LC module. Again, we want to extract general conclusions for each type of circuit, and therefore, for each circuit we explore the space of susceptibilities while monitoring the associated SNR's.

In Fig.(3.14) we plot the result of calculating SNR's as a a function of FS for FB circuits. Here, the input frequency is fixed to the one where the amplitude of the oscillatory response is maximal -although the behavior observed for other frequencies does not change-. In addition, also noise strengths ($D_{II} = D_{SS} = D_{OO}$) are fixed, allowing for a reasonable comparison between different module architectures. When we focus on FM detection we first notice that there is a broad range of possible SNR_{freq}'s for each FS Fig.(3.14A-B). This

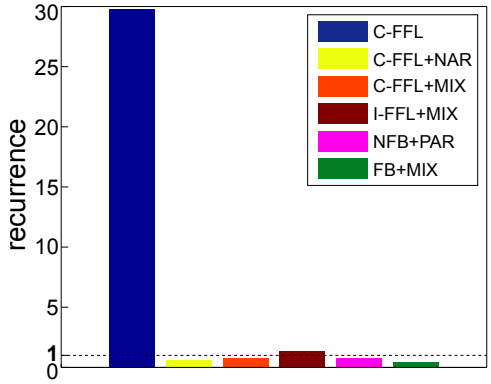


Figure 3.13.: Modules with better AM and FM detection than LC's: statistical analysis of the structure of the modules capable of showing higher s_O and BW than LC's. Y-axes units represent the occurrence of each module type with respect to a distribution of randomly generated structures. NAR and PAR respectively state for negative and positive autoregulations. MIX indicates a mixture of positive (negative) FB's and negative (positive) autoregulations.

is caused by the fact that the different terms in the spectrum are tuned by the individual susceptibilities Eq.(3.28). Unlike for propagation of noiseless FM signals, we can not find a parameter uniquely determining SNR_{freq} as a function of the strength of the interactions and the degradation rates. Now, looking at how the SNR_{freq} range changes when FS increases (starting with a LC configuration, FS=0), we notice that the maximum SNR_{freq} achievable with a FB configuration never exceeds that of the LC. Nevertheless, we notice that, unlike P-FB's, N-FB's reduce the range of possible SNR_{freq}'s, compressing it close to the maximum possible value Fig.(3.14-A). This makes this type of circuits robust transmitters of FM signals. Looking back at the FS regime in which noiseless signals were well transmitted, we see that when working within their good FM transmission regime -large bandwidth-, N-FB's are also able to improve SNR in the frequency domain Fig.(3.14-A). This is thanks to the fact that $A^2(\omega_I)$ increases faster than $P_{fluc}(\omega_I)$ with FS.

Regarding amplitude detection, it is clear that a high output susceptibility would not only amplify signal but also fluctuations [108]. However, s_O and σ_O do not depend on FS in the same way, and therefore, SNR_{amp} evolves differently depending on the module topology. For positive feedbacks, SNR_{amp} increases as a function of FS, since fluctuations are less amplified than signal [108] Fig.(3.14-D). On the other hand, negative feedbacks follow the opposite tendency: a decrease in susceptibility is not followed by an effective noise reduction, and correspondingly SNR_{amp} decreases Fig.(3.14-C). Therefore, for FB structures we see a behavior which is analogous of that described for noiseless signals: N-FB's are robust transmitters of FM signals, while P-FB's work better filtering noise and in AM signals. Moreover, for P-FB's the region in which SNR_{amp} matches the region in which they best transmit noiseless AM signals, in analogy with what happens with N-FB's in FM propagation.

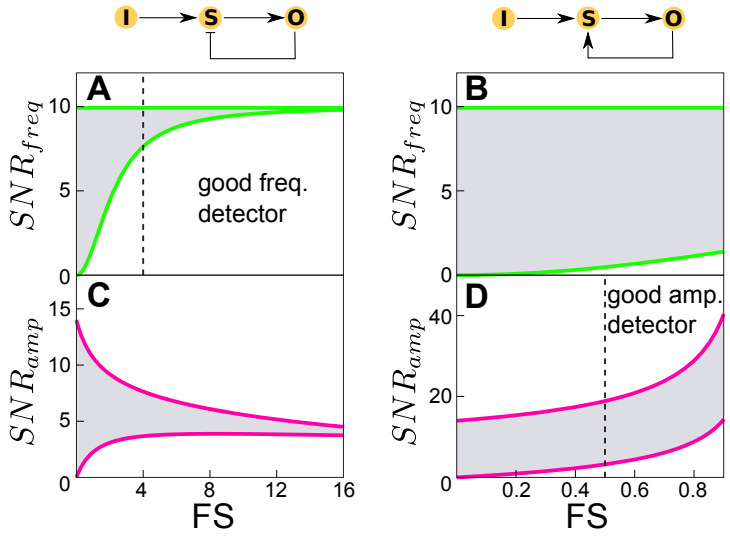


Figure 3.14.: Feedbacks SNR's: in general, a broad range of SNR's can be a chieved for a single FS (grey regions). However, this range is more or less constrained depending on the circuit topology. Green and pink lines respectively indicate SNR_{freq} and SNR_{amp}'s limits. Dashed black lines indicate the start of the good frequency (amplitude) detector regime.

Now we focus the analysis on FFL's. In the previous section we saw that C-FFL's are capable of improving both AM and FM detection. To see how noise affects this properties, we plot again SNR_{amp} and SNR_{freq} as a function of FS for C1-FFL's of random interaction strengths Fig.(3.15). What we see is that both quantities easily increase with respect to LC's, and therefore, both AM and FM detection improve. If we pay attention to the distribution of the different random C1-FFL's in the FS-SNR space, we notice that most of them lie close to the maximum SNR's, indicating that the ability of such structures to propagate both types of signals in a feasible way does not depend on a fine tuning of the parameters. To check all these results, we proceed in a similar way as in Fig.(3.13): we now scan every possible circuit architecture within a broad range of pairwise susceptibilities to find out which topologies are able to improve both SNR_{amp} and SNR_{freq} with respect to the linear cascade Fig.(3.16). What we find is that almost all the structures capable of achieving this goal consist on C-FFL's, although sometimes combined with other interactions.

To study ability of I-FFL's to deal with noisy signals, we take a rather different approach than for the other circuits, combining the analysis with the study of other properties of N-FB's as well. In the previous section we have seen that I-FFL's are capable of operating as a perfect high-pass filters when propagating FM signals. With this in mind, we study the ability of this particular structure to filter noise in the frequency domain. Assuming that in FM signals information is encoded in the temporal behavior of the input species, if this information is encoded in a different time scale than noise, it would be straight forward to separate signal from background. We can translate this idea in terms of bandwidths: on the

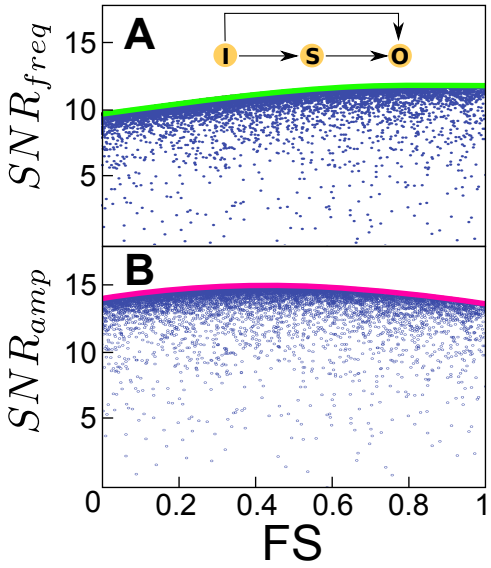


Figure 3.15.: Noisy signal detection in C-FFL's: assigning random pairwise susceptibilities generates circuits of different FS's and SNR's. The maximum SNR's are constrained (respectively green and pink for frequency and amplitude SNR's). However, most of the randomly generated structures lie close to these maximum values (blue dots).

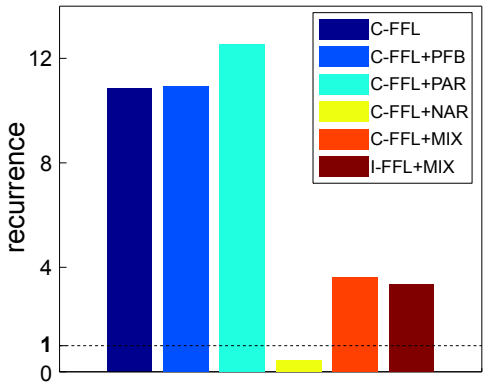


Figure 3.16.: Modules with better SNR's than LC's: statistical analysis of the structure of the modules capable of showing higher SNR_{amp} and SNR_{freq} than LC's. Y-axes units represent the occurrence of each module type with respect to a distribution of randomly generated structures. NAR and PAR respectively state for negative and positive autoregulations. $MICX$ indicates a mixture of positive (negative) FB's and negative (positive) autoregulations.

one hand we have the bandwidth of the propagated oscillations (BW_{osc}), which correspond to the bandwidth of the information transmitted along the network. On the other hand, the fluctuations power spectra has its own bandwidth (BW_{fluc}), which summarizes the frequency range in which noise is still relevant Fig.(3.17).

In the *Methods* section we already showed that, the frequency-dependent term describing the propagation of the input intrinsic noise in Eq.(3.28) is the same as the one describing

oscillations propagation Eq.(3.35). Therefore, if no intrinsic noise is present in the sensor and output species -$D_{SS} = D_{OO} = 0$-, the bandwidth of both fluctuations and propagated oscillations would be the same. Accordingly, if intrinsic noise is present, then both bandwidths differ depending on the circuit topology.

Carefully looking at Eq.(3.28), we notice that there are two terms potentially acting as high-pass filters (numerators with ω^2). The first one is in the term representing propagated fluctuations from the sensor species to the output. This kind of high-pass filter is only present if a N-FB interaction exists. Therefore, within some parameter range, N-FB's will present faster fluctuations than oscillations, allowing for a noise-free frequency regime. This is shown in Fig.(3.17). At relatively high FS's, there are negative feedback circuits whose BW_{fluc} is always at higher frequencies than BW_{osc}, allowing for a frequency window in which information can be robustly transmitted. The second high-pass filter term is related with the propagated fluctuations from the signal to the output, when a FFL interaction is active. This filter is also present in the response to propagated oscillations, and dominates in I-FFL's Fig.(3.17). While this term allows for propagation of faster signals, the two additional terms in Eq. (3.28), which are related to intrinsic circuit fluctuations in sensor and output species, model low-pass filters. Their effect is a shift of BW_{fluc} to lower frequencies. Therefore, I-FFL's may act as transmitters of fast FM signals in which slower noise is filtered. This is shown in Fig.(3.17). If the direct susceptibility s_{OI} is larger than the global susceptibility s_O -$FS > 1$-, the high-pass filter for the transmitted oscillations dominates, leading to a frequency window at high frequencies in which noise is already low but oscillations are still well transmitted. Note that in this case the bandwidth of the fluctuations spectrum is still large, and no perfect filter exists, unlike for the negative feedback case.

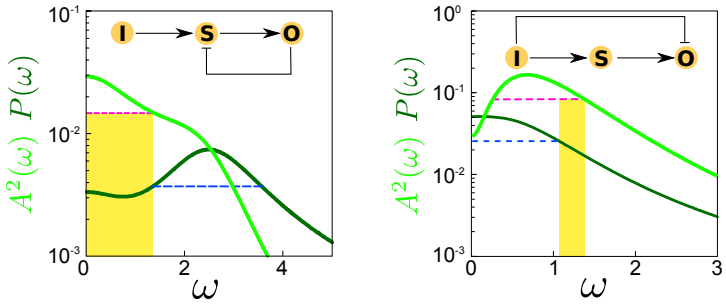


Figure 3.17.: FM noise filtering in a N-FB and an I-FFL: oscillations amplitude $A^2(\omega)$ (light green) and fluctuations power spectra $P(\omega)$ (dark green) are represented together. BW_{amp} and BW_{freq} are respectively represented by pink and blue dashed lines. Regions of feasible FM signal transmission are delimited in yellow.

To conclude, we see that N-FB's act as good noise filters for low-frequency FM inputs, while I-FFL's perform this task when the frequency of the FM signals is high.

Validating susceptibility ranges with experimental data

In the last sections we have reached some conclusions on the signal propagation abilities of diverse genetic modules. Most of the AM and FM signal propagation and noise filtering properties described in this study, like the existence of high-pass filtering regimes in I-FFL's or the SNR increase in C-FFL's with respect to LC's, depend at some extent on the specific regime the modules are operating in. Are real systems capable of achieving these abilities under natural conditions? To address this question we analyze experimental data provided by Kaplan et al. [123, 124]. They measured the *in-vivo* production rate of different genes involved in sugar metabolism in *E. coli*. These genes are mainly arranged forming coherent and incoherent FFL modules which are specially interesting to complete our study.

We analyze two of the modules presented by Kaplan and co-workers. In both modules, the second messenger cyclic adenosine monophosphate (cAMP) activates the cAMP receptor protein (CRP), which works as a transcription factor regulating downstream genes. The first module consists of an I-FFL structure controlling expression of genes in the gal operon. In this case, CRP positively triggers the activity of GalS and GalE. At the same time, GalE is repressed by GalS, forming the I-FFL circuit. Apart from cAMP, D-Galactose, which activates GalS, is the other input of the system Fig.(3.19). The second module is arranged in a C-FFL structure and regulates expression of some genes in the mal operon: CRP activates MalT and MalE. Meanwhile, MalE is also activated by MalT. Maltotriose post-transcriptionally activates MalT by favouring MalT activation by self-association Fig.(3.20) [125].

For both modules, the input functions (production rates) of GalE and MalE are measured as a function of cAMP and respectively Galactose and Maltotriose concentration. The first system is modeled by the authors with the following ODE system

$$\begin{aligned} \frac{dC}{dt} &= \alpha_C[cAMP] - \delta C \\ \frac{dG_S}{dt} &= \frac{C^{h_1}}{1+[galactose]/k_g} - \delta G_S \\ \frac{dG_E}{dt} &= \frac{1}{1+(G_S/k_{G_S})^{h_2}} \cdot \frac{(C/k_C)^{h_3}}{1+(C/k_C)^{h_3}} - \delta G_E \end{aligned} \tag{3.46}$$

Parameters for this first model were estimated by Kaplan et al., leading to values of the constants which allow the model to be in good agreement with the experimental data: $k_C = k_{G_S} = k_g = 5\ mM$, $h_1 = h_2 = h_3 = 1$. Degradation is assumed to take around one cell cycle, and therefore, $\delta = log(2)/\tau_{div}$, being $\tau_{div} = 40\ min$, while $\alpha_C = \delta$, in such a way that CRP at equilibrium is the same as cAMP concentration.

Similarly to the previous system, we model the maltotriose module as follows: MalT binds promoters of the malEFG, malPQ and malK operons as an oligomer stabilized by the inducer sugar [22, 126]. CRP and MalT both act cooperatively to activate MalE and MalK production

[125]. With these facts in mind, we write down a model of a C-FFL based on Hill type regulation functions to fit the experimental production rates of Kaplan's experiments

$$\begin{aligned}\frac{dC}{dt} &= \alpha_C[cAMP] - \delta C \\ \frac{dM_T}{dt} &= \frac{\alpha_T + \beta(C/K_1)^{h_1}}{1+(C/K_1)^{h_1}} - \delta M_T \\ \frac{dM_E}{dt} &= \alpha_E \frac{s(C/K_2)^{h_2}(C/K_3)^{h_3}}{1+s(C/K_2)^{h_2}(C/K_3)^{h_3}} \cdot -\delta M_E \end{aligned} \tag{3.47}$$

Individually fitting the production rates of MalT and MalE to their corresponding Hill functions in the model allows us to determine the parameters which better fit the model to the experimental data: $K_1 = 5.6\ mM^{-1}$, $K_2 = 7.6\ mM^{-1}$, $K_3 = 7.1\ mM^{-1}$, $h_1 = 1.5$, $h_2 = 2.4$, $h_3 = 1.3$. From the experimental observations for the maximum MalT and MalE production rates, we took $\beta = 65\ mM$ and $\alpha_E = 50\ mM$. We allowed for a small basal activation of MalT by CRP $\alpha_T = 4.4\ mM$. The model is in good agreement with the experimental data, being able to reproduce the system's behavior at equilibrium for different concentrations of sugar and cAMP Fig.(3.18).

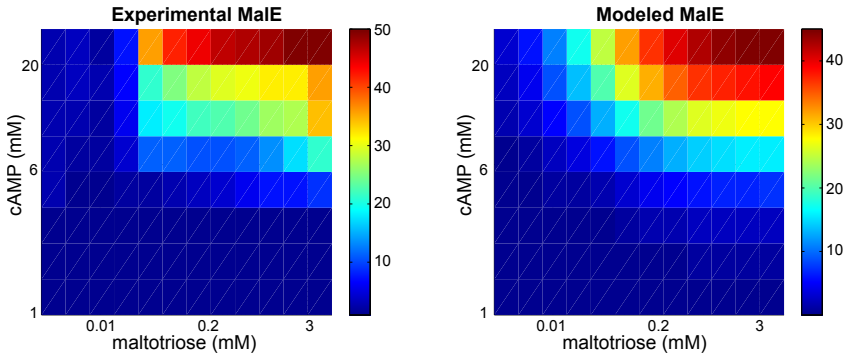


Figure 3.18.: Experimental and modeled MalE response

Once the models are validated, we can use them to compute pairwise susceptibilities which, in turn, allow us to characterize the AM and FM signal detection properties of these real modules by using our theoretical predictions. We focus first on the I-FFL circuit. For different galactose concentrations, the GalE response shows a maximum Fig.(3.19) [124], revealing the existence of an optimal cAMP concentration for AM transmission. This amplitude-filter behavior was also observed in synthetic incoherent FFL circuits in E. coli [22, 126]. The reason behind this is that the output susceptibility s_O changes sign. This occurs because the direct CRP-GalE activation saturates and the negative interaction starts dominating, repressing the GalE promoter at high cAMP levels. Moreover, this implies that FS is high around the maximum of the input function, and therefore we expect band–pass filtering for FM signals. This is shown in Fig.(3.19), where the bandwidth of the GalE response for an oscillatory cAMP input is plotted in as a function of cAMP and galactose concentrations. White lines

indicate the cAMP boundaries where the system behaves as a band–pass frequency detector. Moreover, taking into account the noise in biochemical reactions, we observe that this system is also able to filter fluctuations for high–frequency oscillations by the mechanism discussed in the previous section (noise filter range is marked with black solid lines in Fig.(3.19)).

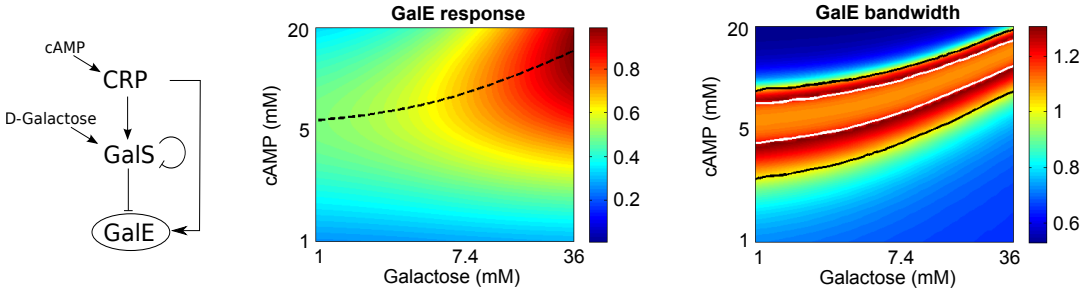


Figure 3.19.: Properties of the GalE I-FFL system: apart from the I-FFL structure, the circuit has a negative autoregulation in the GalS component. This interaction is neglected in the model because its effect is small. Black dashed line indicates the rim where GalE expression is maximum for each galactose concentration. White lines in the bandwidth plot delimit the region in which the circuit behaves as a band-pass filter. Solid black lines state for the region in which the condition represented in Fig.(3.17) is satisfied.

Now we analyze the C-FFL system. If we compare it with the LC's limit, where we neglect the CRP-MalE interaction, we corroborate that both bandwidth and output susceptibility are larger when the extra interaction of the C-FFL configuration is present within the whole input range Fig(3.20). One of the main results is that, for moderate s_{OI}'s, a C-FFL is capable of improving signal detection in the presence of noise, giving SNR's beyond the linear cascade limit. We plotted the SNR_{amp} and SNR_{freq} divided by the maximum value achieved by a linear cascade module in the whole input range Fig.(3.20) -keeping identical susceptibilities for CRP-MalT and MalT-MalE interactions-. As seen from the plotted data, SNR can improve up to factor of two (AM) or three (FM).

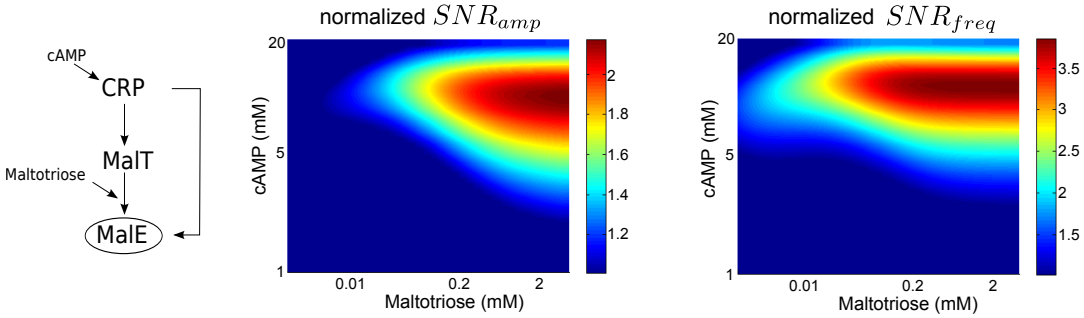


Figure 3.20.: Properties of the MalE C-FFL system: SNR's are normalized by their maximum values in the LC configuration, where the CRP-MalE interaction is neglected.

The study of this systems confirms that gene modules in their natural environment are

able to operate in regimes in which the signal detection and noise filtering properties we have unveiled in the previous sections are in fact present. These result suggest new wet-lab experiments to test our predictions, which could be possibly performed by means of microfluidic devices. As described in chapter 2, these devices are able to stimulate cells with oscillatory signals of well controlled frequency, allowing for good estimates of the modules bandwidths.

3.3.2. Correlations between structure, dynamics and signal propagation

So far we have systematically characterized the main signal propagation and noise filtering abilities of simple gene circuits. In our analysis we have mainly focused on the most abundant interactions found in gene networks: feedback and feedforward interactions. In the case of feedforward modules we have confirmed that, in fact, this type of interaction is essential when specific signal transmission and noise filtering abilities are to be achieved Figs.(3.13,3.16). Although our analysis provides valuable information about the effect of each interaction on overall signal transmission, it does not provide a description of the dynamical response of the modules and how it is related with the signal propagation properties at equilibrium (transmission of AM and FM signals, and noise propagation). Thus, to partly address this question we now focus on how the dynamic response of the circuits depends on their structure, and how signal propagation arises from this dynamical behavior.

To characterize the structure of the circuits we have already introduced the pairwise susceptibilities, which account for the interactions between the different circuit species. In addition to individual susceptibilities, we have also defined some combinations of them, like feedback and feedforward strengths, as relevant structural features. As steady state responses we use the already defined bandwidth of the propagated oscillations Eq.(3.35) and the coefficient of variation of the propagated noise, third term in Eq.(3.21). Finally, to account for the dynamic response of the circuits we now define the response times, T_{on} and T_{off}, as the time spent by the output to travel half the way to its new equilibrium concentration, respectively following an increase or a decrease in the input concentration Fig(3.21). Similarly to FS and s_O, together with T_{on} and T_{off} we also define as dynamical features two combinations of them: their sum $\Sigma T = T_{on} + T_{off}$, and their difference $\Delta T = T_{on} - T_{off}$.

Apart from providing an improved understanding of the signal transmission properties of gene networks, the systematic study of the dynamic responses of gene circuits is by itself of great importance to understand how gene networks work. Previous studies have suggested the dynamics of gene networks to be important in determining their actual topology [127], as well as in encoding signaling information [128]. The relationship between the topology of gene modules and their dynamical behavior has been investigated in previous works [129–131]. Closely related to the results we present here is the work by Alon and coworkers in which they characterize the response times of circuits with feedback and feedforward

interactions under different cis-regulatory functions [129, 132]. Moreover, the influence of dynamical features on the steady state response of the circuits has already been addressed in previous works. For example, difference in response times is described to be involved in fluctuations propagation in feedback systems [133].

Following all these ideas, the analysis we have performed over relevant circuit structures Fig.(3.3) allows us to unveil the connections between the structure of the circuits, their various steady state responses to signals of diverse characteristics, and the dynamics of their response to a change in the input.

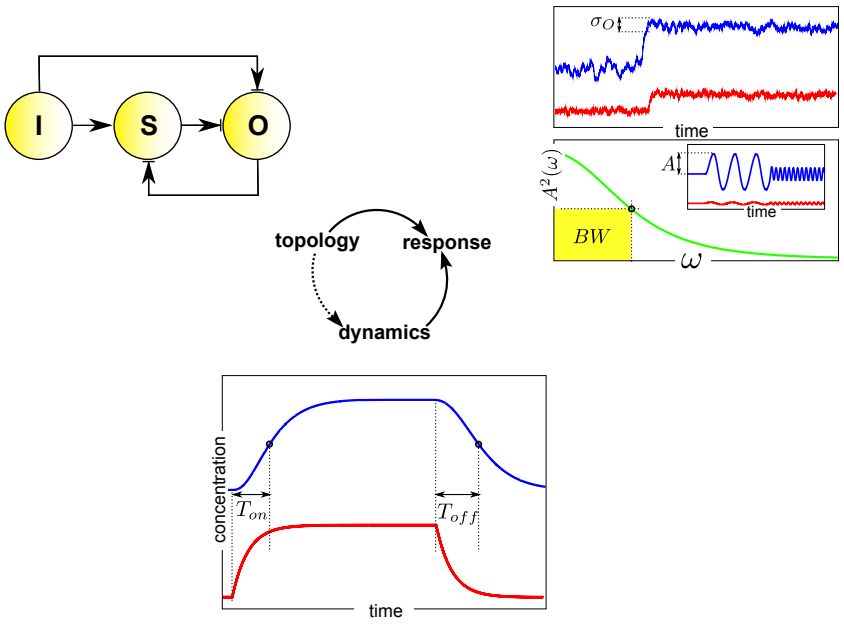


Figure 3.21.: Relations between structure, dynamics and response is represented in the central circle. It is studied for the circuits represented in Fig.(3.3). Response times, T_{on} and T_{off} are represented in the bottom plot. Red and blue lines respectively state for the input (I) and the output (O) concentration. Responses of the circuits are measured in terms of the bandwidth of a propagated oscillation BW and the coefficient of variation of fluctuations in the output propagated from the input σ_O.

Structure and dynamics determine signal propagation

Circuits with the very same topology and biochemical components, when embedded in larger regulatory or signaling networks, can differ substantially in interaction strengths and dynamic response, due for instance to differences in cell cycle state or signaling context [134]. To deal with the variability imposed by these differences in conditions, we sample the space of susceptibilities of the three component networks presented in Fig.(3.3), with both 'AND' and 'OR' regulatory logic -see Table(3.1)-, and use the theoretical approximations Eqs.(3.35,3.28,3.21) to calculate the response properties. Depending on the network and its logic gate, each structural feature constrains the space of possible responses in a different

way. This is illustrated in Fig.(3.22) for a C1-FFL. Uniformly sampling individual susceptibilities s_{ij}, we see that some structural features restrict more than others the space of possible responses. For instance, the bandwidth of the propagated oscillations is specially constrained by the sensor-output interaction s_{OS}, showing a clear anticorrelation, while the direct input-sensor interaction s_{OI}, exhibits a weak positive correlation and allows for a wider response range. As expected from the previous results, the FS fully constrains the bandwidth, exhibiting a one-to-one relation between both quantities. On the other hand, noise propagated from the input to the output increases with both s_{OS} and s_{OI}, although ultimately the output's susceptibility s_O is the dynamical feature which better determines noise propagation.

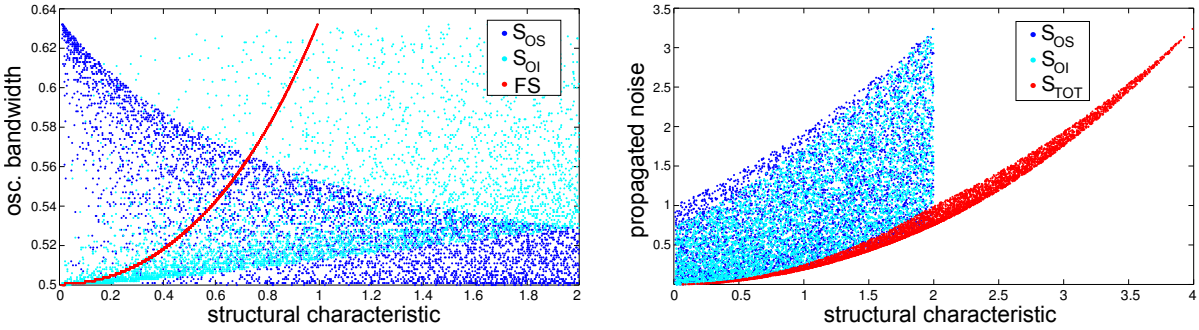


Figure 3.22.: Dependence of bandwidth and propagated noise with respect to circuit structure a class 1 coherent feedforward loop under 'AND' logic. Calculations have been performed fixed $s_{SI} = 1$. s_{OS} and s_{OI} are randomly varied between 0 and 2.

To systematically quantify the degree of constraint imposed by the different structural features on the output response, we compute the correlations between both variables. We use two correlation measures: the Spearman's rank correlation coefficient (SC) and the mutual information coefficient (MIC), both already described in the *Methods* section. Mutual information quantifies how much knowledge about one variable we gain by knowing the other, and it has been used in similar contexts to study form-function relations in biological networks [120], as well as to detect constraints between different sensing characteristics in allosteric models of transcription factors [121]. It has the advantage that is able to account for non-monotonic dependencies between variables: for instance, given two variables X and Y, if their dependence is of the form $Y = sin(X)$, we would get MIC=1, while Spearman's rank correlation would be SC=0. On the other hand, MIC does not provide information on the sign of the correlation: given $Y = X$ or $Y = -X$, we would always get MIC=1, while SC would be respectively 1 and -1. At the same time, the value of MIC is strongly influenced by the amount of information available in our data: if we only characterize the behavior of Y for a limited range of X's, there would be a high degree of uncertainty on the 'real' dependence of both variables, and therefore, low MIC values would be computed. Thus, even if there is a one-to-one dependence between the variables, MIC will be low if the sampling range is small. Therefore, we see that both variables are somehow complementary: while SC

provides information on the sign of the correlation, MIC is able to deal with non-monotonic dependences. Since in our analysis we would in principle expect any possible dependence, either monotonic or not, increasing or decreasing, it seems reasonable to have two different methods to compute correlations in the less biased way.

We compute correlations between structural features and responses for all the different circuits described in Fig.(3.3), unveiling general structural constraints on the different responses. This very same procedure can be also applied to elucidate constraints in the response imposed by the dynamical features of the circuits, therefore helping to complete the circle depicted in Fig.(3.21). Given a specific circuit topology, the strength of its interactions determines both signal propagation and dynamics. Therefore, sampling the structural space by randomly assigning susceptibilities to all the interactions generates a set of different dynamical features and output responses over which correlations can be calculated.

Fig.(3.23A,B) shows the results for a C1-FFL, with empty bars representing SC and solid bars MIC. Bandwidth correlations are presented in Fig.(3.23A). Looking at SC, both activation and deactivation times are strongly anti-correlated with the bandwidth. As we would intuitively expect, the faster the response (the lower the response times), the higher the frequency (bandwidth) the network can propagate. Accordingly, the sum of both response times ΣT shows even stronger anti-correlation than the individual ones ($SC \simeq -1$). Although it does not give us information about the sign of the correlation, MIC corroborates this result, showing even a bigger difference between correlations of individual response times and ΣT. On the structural side, different correlation signs are found for different interactions: on the one hand, both s_{SI} and s_{OS} show a relatively high anti-correlation. These are the interactions of the linear cascade branch of the FFL, therefore the linear cascade susceptibility ($s_{LC} = s_{SI} \cdot s_{OS}$) exhibits high anti-correlation too. On the other hand, the direct interaction s_{OI} shows a clear positive correlation. This results can be easily explained: since more species are involved in its signal propagation, the linear cascade's branch represents the slow input-output path of the circuit. Unlike this branch, the direct interaction s_{OI} only involves the input and the output themselves, representing the circuit fast route. Therefore, the stronger the linear cascade branch the slower the circuit and the lower its threshold frequency. On the other hand, the stronger the direct interaction the faster the circuit and the higher the threshold frequency. These opposite behaviors are condensed in one quantity which uniquely determines the network bandwidth: as we already showed when the trade-offs between AM and FM signal propagation were described and in Fig.(3.22), FS exhibits a one to one relation with the bandwidth. This result is confirmed here, where the correlation between FS and bandwidth is found to be $SC \simeq 1$, being this result provided by both SC and MIC.

Fig.(3.23B) shows the results for noise propagation. From the dynamical point of view the main result is that the difference between activation and deactivation times ΔT shows a strong anti-correlation. This result agrees with the previously mentioned result presented in [133] for feedback systems. On the structural side, the total susceptibility of the network

plays the main role in noise propagation, something that has been already shown [108]. This result was already stated in Fig.(3.22B), where points representing propagated noise and total susceptibility relations are distributed in a very narrow area. Here, both high SC and MIC for s_0 quantify this little dispersion.

All the previous results correspond to a type 1 coherent FFL with 'AND' logic. In principle, we would expect circuits with different topologies to exhibit different tendencies, with different topological and dynamical constraints on the output response. To get a better insight on this question, we extend the previous analysis to different topologies and see if there are general dynamical and structural characteristics constraining the different responses under analysis. Fig.(3.23C,D) summarize these results for all the networks under study, with both 'AND' and 'OR' logic gates. Displayed correlations correspond to absolute values of SC, although similar results are obtained for MIC. Reading these matrices from top to bottom helps us to visualize general correlations between response characteristics and all structural/dynamical features. Bandwidth correlations are displayed in Fig.(3.23C). We notice that the propagation of oscillatory signals is, as expected, uniquely determined by FS (here, both feedforward or feedback strength). On the dynamical side, the results extracted from Fig.(3.23A) are general for every single network and logic gate: ΣT is the main dynamical feature involved in the ability of all the circuits to propagate oscillatory signals. Propagated noise correlations are displayed in Fig.(3.23D). Again, on the structural side we find that total susceptibility plays a key role in noise propagation. This is directly related with the assumption made in the previous section in which we stated that transmission of AM signals is directly characterized by the output susceptibility: the higher s_O the stronger the output response to changes in the input. Therefore, fluctuations propagated from the input to the output -with no intrinsic noise in the sensor or the output themselves- are amplified proportionally to that magnitude. Unlike for the bandwidth, we cannot find general dynamical features clearly linked with noise propagation. ΔT shows strong correlations for FFL's under 'AND' logic, but this is not the case for 'OR' gates. Looking at the linear approximations derived to compute response times Eqs.(B.13,B.15), we notice that activation and deactivation times depend on the susceptibilities and equilibrium concentrations of the 'off' and 'on' states respectively. Therefore, unlike for noise propagation or bandwidth, the specifics of the kinetic model have a strong influence on the dynamics of the system, ultimately determining the differences between 'AND' and 'OR' logic circuits. Although differences in response times was presented as a key dynamical feature influencing noise propagation in feedback systems [133], the correlation between both quantities is low when the system is mapped for random interaction strengths. This means that, although some influence may be present, the overall constraint that this specific dynamical feature induces on propagated noise on such systems is low.

To summarize, we could say that, for every network under study, we have found that feedback or feedforward strengths uniquely determine the ability of the circuit to propagate oscillatory signals. In addition, the output susceptibility s_O strongly determines the amplitude

of propagated noise, as expected from the assumption that propagation of AM signals is directly linked with that magnitude. At the same time, total response time ΣT and oscillations propagation are, as intuitively expected, features which are strongly correlated. Nevertheless, although response time difference ΔT seems to be the best candidate, no general dynamical feature is found correlated with noise propagation.

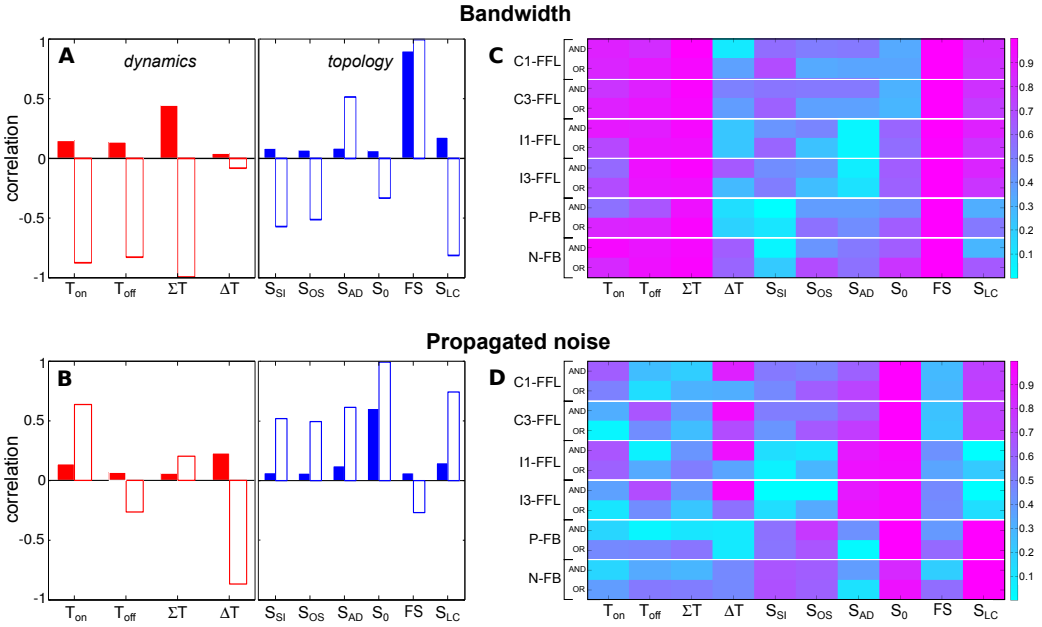


Figure 3.23.: Correlations between response features and different topological and dynamical features. $s_{LC} = s_{SO} \cdot s_{OS}$ is the susceptibility of the linear cascade part of the circuits. s_{AD} represents the additional susceptibility: s_{SO} for feedbacks and s_{OI} for feedforwards. For a C1-FFL under 'AND' logic, correlations between dynamical(red)/topological(blue) features and response characteristics are computed: bandwidth of propagated oscillations (A) and propagated noise (B). Both Spearman's rank correlation (empty bars) and mutual information coefficients (filled bars) are included. For all the analyzed circuits and logic gates, absolute value of Spearman's rank correlation are computed between dynamical/topological features and response characteristics: bandwidth of propagated oscillations (C) and propagated noise (D).

Structure determines dynamics

The topology of a circuit, together with the strength of its interactions, determines its reaction to input signals. When we talk about that reaction we can state for its speed (response times), its amplitude (proportional to the susceptibility [115]) or for the propagation of complex signals (oscillatory signals or fluctuations). It is clear that both the speed and the amplitude of the reaction will ultimately determine the ability of the circuit to propagate signals. For example, one would expect a fast circuit (with low response times) to be good in propa-

gating oscillations (high bandwidth). In the previous section we have shown that there are, in general, key structural and dynamical features strongly involved in the ability of the circuits to respond to specific signals. These links correspond to the solid lines of the circle in Fig.(3.21). To close that circle, we now focus on the study of the relations between the different topological features and the dynamics of the networks, (dashed line in Fig.(3.21)).

To address this question we use a different approach than in the previous section: first we compute correlations between all dynamical and structural properties of the circuits to build the adjacency matrix shown in Fig.(3.25A). Afterwards, we perform a principal component analysis (PCA) over this matrix in order to unveil the connection between different features Fig.(3.25B).

Fig.(3.25A) shows the adjacency matrix calculated for a C1-FFL. Absolute values of SC are calculated to quantify correlations. From there we can see that, as expected for this specific circuit, there is a high correlation between ΣT and FS, as well as between δT and s_O. Therefore, in the principal component space ΣT is close to FS and ΔT to s_0, Fig.(3.25B).

Finally, Figs.(3.25C-D) summarize the results of this analysis for all the circuits and logic gates under study. In these matrices we show the distances in the principal component space between key dynamical features, like ΣT and ΔT, and all the topological features. Each distance d_{xy} between features x, y is computed taking into account as many principal components ($PC1, PC2, PC3$,...) as necessary to explain at least 85% of the observed variability Fig(3.24)

$$d_{xy} = \sqrt{(x_{PC1} - y_{PC1})^2 + (x_{PC2} - y_{PC2})^2 + (x_{PC3} - y_{PC3})^2 + ...}$$

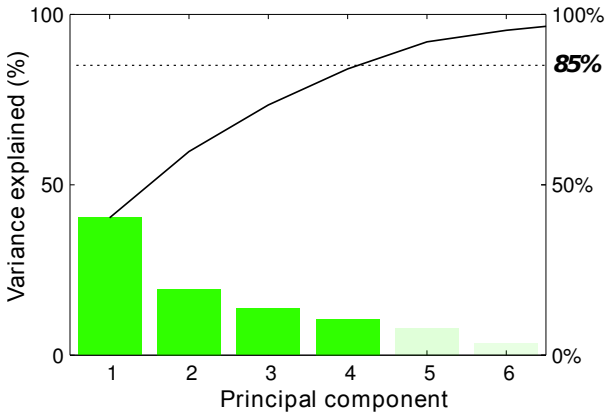


Figure 3.24.: Weight of the different principal components in a P-FB circuit: to calculate distances in the principal component space, we take as many principal components as necessary to explain at least 85% of the observed variability (black line). In this particular case, 4 principal components are taken into account.

Fig.(3.25C) represents the distance between ΣT and the different structural features. In the previous section we showed that FS uniquely determines the ability of the circuits to prop-

agate an oscillatory signal. At the same time, the propagation of these signals was strongly linked to the 'speed' of the circuit, quantified by the time ΣT it spends to respond to a transient pulse. Here we close the circle by showing a general relation between ΣT and FS: the distance between ΣT and FS in Fig.(3.25C) is close to zero for every single motif and logic gate under study. Therefore, FS is a topological feature which is strongly involved in determining both the ability of the circuit to propagate oscillations and its total response time.

Fig.(3.25D) represents the distance between ΔT and the different structural features. In the previous section we found the total susceptibility s_0 to be a general structural characteristic determining noise propagation, but we could not find a general dynamical feature involved in this process. The best dynamical candidate to play this role seemed to be ΔT, but its influence in noise propagation was not general for all the circuits and logic gates. Therefore, when we calculate distances in the principal component space between ΔT and all the different susceptibilities and FS, we do not find a general rule. s_0 and ΔT are close to each other for FFL's with 'AND' logic gates, which are the same circuits which showed a correlation between ΔT and propagated noise in the previous section. In addition, s_0 and ΔT are also close for type 3 coherent and incoherent FFL's with 'OR' logic and for negative FB with an 'AND' gate. Basically, the pattern we see in the s_0 column of Fig.(3.25D) is the same that the pattern along the ΔT column of Fig.(3.23D). Therefore, we conclude that the lack of correlation between ΔT and σ is ultimately determined by the lack of correlation between s_0 and ΔT. In analogy to AM transmission, s_0 is always strongly involved in noise propagation, so a $s_0 \rightarrow \sigma$ link always exists. At the same time, there are some circuits in which ΔT is highly correlated with σ too: these very same circuits where the $\Delta T \rightarrow \sigma$ link exists, show also a connection $s_0 \rightarrow \Delta T$.

To give a brief summary, in this section we have found that, if there is one link $topology_1 \rightarrow response_1$ and another $dynamics_1 \rightarrow response_1$, then there is always a $topology_1 \rightarrow dynamics_1$ link too. Here the sub-index 1 indicates a specific topological, dynamical and response feature.

3.4. Discussion

In this third chapter of the thesis we have introduced a framework to study signal propagation and dynamical characteristics of simple gene modules. The derived formalism consists of analytical expressions to quantify both AM and FM signal propagation, as well as some noise properties (amplitude and frequency content) and the dynamic response of the circuits (response times). All these expressions are written in terms of the same biologically relevant magnitudes: strength of the interactions between the network species (susceptibilities and elasticities) and their degradation rates. Although all the theoretical expressions depend on

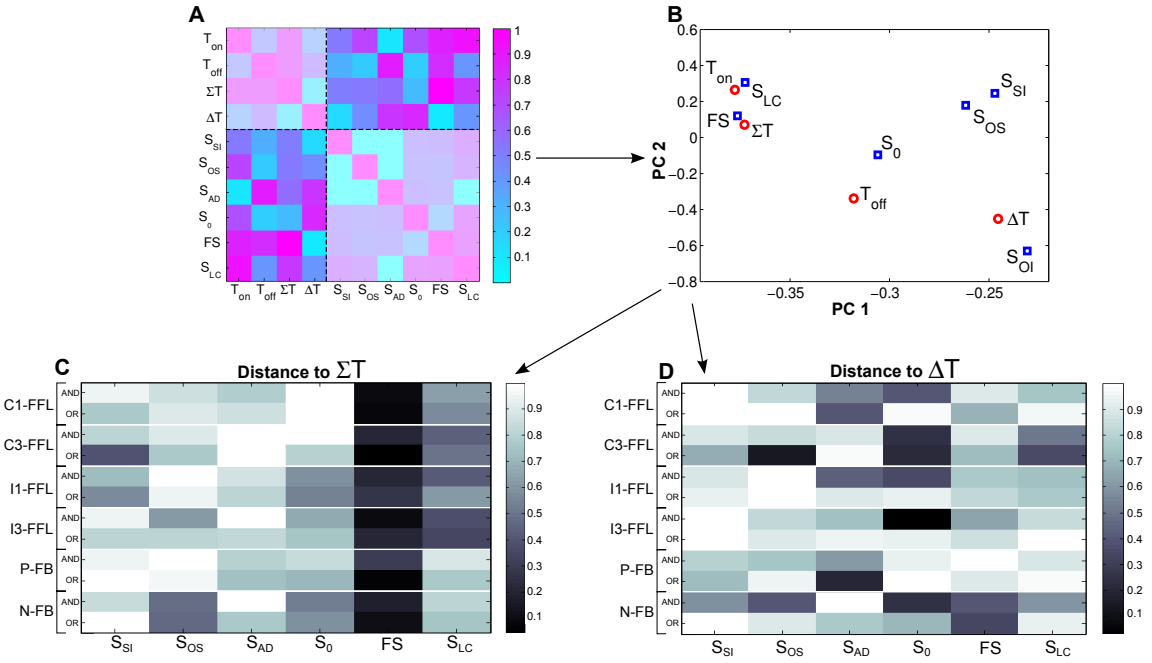


Figure 3.25.: Correlations between topological and dynamical features: $s_{LC} = s_{SO} \cdot s_{OS}$ is the susceptibility of the linear cascade part of the circuits. s_{AD} represents the additional susceptibility: s_{SO} for feedbacks and s_{OI} for feedforwards. A: dynamical and topological features adjacency matrix calculated as the absolute value of Spearman's rank correlation coefficient for a class 1 C-FFL under 'AND' logic. B: principal component (PC) analysis based on the matrix plotted in A. C and D: generalization for all the circuits and logic gates of the results shown in A and B respectively, addressed by calculating distances in the PC space from ΣT and ΔT to the rest of the dynamic and topological features.

linear approximations around steady states, they seem to make good predictions on the behavior of the circuits, even for strongly non-linear interactions Fig.(3.7-3.8). Ultimately, this formalism is used to extract general conclusions on the response features of simple three component modules.

To characterize the signal propagation features which are more commonly associated with specific circuit topologies, we have performed a statistical analysis in which the behavior of the modules is quantified for random sets of interaction strengths. This allows us to get a general picture of the response ranges in which different topologies lie.

We started studying propagation of both AM and FM signals, first from a deterministic point of view and then in the presence of random fluctuations on species production. We focused mainly on four types of circuits which are commonly found in real gene networks: positive and negative FB's and coherent and incoherent FFL's, and compared the influence that their extra interactions produce in a simple 3 component linear cascade. AM and FM

detection are independent in this simple LC circuit: while FM detection is constant along the whole range of interaction strengths, FM detection is governed by the length of the cascade [122] and its slowest decay rate Fig.(3.10) [109].

We first focused in propagation of noiseless signals. For FB circuits, we were able to find a biologically relevant quantity which uniquely determines FM signal propagation: a combination of pairwise susceptibilities named feedback strength. The bandwidth of the propagated oscillations is solely given by this quantity, independently of individual interaction strengths. Signal propagation exhibits a trade-off in FB circuits: while N-FB's are able to improve FM detection compared to LC's, their capability of propagating AM signals is reduced Fig.(3.11). On the contrary, P-FB's behave in the opposite way: they are not good FM detectors, but they improve AM detection thanks to their positive extra interaction Fig.(3.11).

Similarly to FB structures, we found a different combination of pairwise susceptibilities, named feedforward strength, that uniquely determines the bandwidth of propagated oscillations in FFL circuits. The signal propagation properties of these circuits are very different than those of FB's. They do not exhibit this AM-FM trade off: C-FFL's are able to improve both AM and FM detection with respect to the LC configuration Fig.(3.12). This result was further corroborated when we looked for the types of structures able to exhibit better AM and FM detection than LC's Fig.(3.13): here we found that only C-FFL's are significantly able to perform this task. On the other hand, I-FFL's do not improve AM detection, but are the only configuration able to exhibit a clear high-pass filtering regime in FM propagation Fig.(3.12). This has been mentioned in previous studies, both theoretically and experimentally [22, 124, 126, 135].

To quantify signal propagation in the presence of fluctuations, we defined SNR's both for AM and FM signals Eq.(3.10-3.11). Unlike for noiseless signals, no quantity was found determining SNR_{freq}, neither for FB's nor for FFL's. Nevertheless, performing a similar analysis than before, we found that N-FB's never exceed the LC's ability to propagate information encoded in noisy AM or FM signals, although they compress the region of possible SNR_{freq}'s close to the maximum possible value, making them robust FM transmitters Fig.(3.14). On the contrary, P-FB's do improve SNR_{amp} with respect to LC's Fig.(3.14), as previously suggested by Hornung and coworkers [108].

The picture for FFL's is, again, slightly different. C-FFL's are able to improve both SNR_{freq} and SNR_{amp}. Moreover, they seem to do it in a robust manner, since most of the randomly generated C1-FFL's exhibited better SNR's than LC's Fig.(3.15). This is again tested by finding the structures which are able to achieve higher SNR_{freq} and SNR_{amp} than LC's: although usually combined wit other types of interactions, most of the selected circuits have a C-FFL interaction Fig.(3.16). The response to noisy FM signals by I-FFL's is studied in comparison with N-FB's. Both circuits are able to filter noise in the frequency domain by propagating oscillations in a frequency range in which fluctuations are small, but they do it in two different ways: N-FB's are able to work as low-pass filters, filtering fast fluctuations [136] while

allowing propagation of slower signals. On the other hand, I-FFL's can work as high-pass filters, allowing fast signals to propagate while filtering slower fluctuations Fig.(3.17).

Finally, to check that real networks naturally operate in the necessary regimes to exhibit some of the discussed properties, we worked with FFL models presented by Kaplan et al. [123, 124]. The possibility of band-pass filtering and noise-tolerance in a real I-FFL was found for the GalE operon Fig.(3.19). Moreover, the C1-FFL associated to maltose metabolism could exhibit better AM and FM signal detection in the presence of noise (3.20) than a simple LC in which the CRP-MalE interaction is neglected.

In the second part of the *Results* section we have characterized the correlations between the structure of the circuits, their signal propagation abilities and their dynamic response, closing the circle in Fig.(3.21). By doing so we have gained a deeper understanding on the dynamical reasons that allow different circuits to respond differently to diverse input stimuli. Moreover, by unveiling the relationships between circuit topology and the dynamic response, we have provided information which may be useful in the design of synthetic gene circuits [106, 107].

Similarly to the previous results, a statistical approach has been followed in order to get a parameter-independent idea of the influence of each interaction on the dynamics. Principal component analysis performed over mutual information coefficients used to compute correlations between features has been previously used: Mugler and coworkers [120] have examined the correlations between topology and the possible sets of steady state responses in a LC with different FB interactions. Martins et al. [121] have analyzed correlations between sensing characteristics in a model of allosteric sensing.

The most relevant results obtained from our study are obtained by examining from top to bottom the colored matrices in Figs.(3.23-3.25). By studying propagation of FM signals, we can see that the the sum of response times plays a main role in the ability of the circuits to propagate oscillations, irrespectively of the type of circuit and the logic interaction it is operating with. Moreover, as we already demonstrated, the feedforward strength is the topological feature which uniquely determines the bandwidth of propagated oscillations. Accordingly, when performing a principal component analysis, we see that the FS and ΣT are always close in the principal component space, therefore completing the closed picture in which topology determines both signal propagation and dynamic response, which are at the same time strongly correlated. In this case, FS determines both bandwidth and ΣT, while bandwidth and ΣT are themselves strongly correlated.

At the same time, we have tried to obtain similar results when the signal propagation feature is the amplitude of the propagated fluctuations. As expected, the overall susceptibility s_0, which determines the relative change in the output given a change in the input, is strongly correlated with this quantity. Nevertheless, the correlations with dynamical features are not that clear: as suggested by Wang and coworkers [133], the difference between response times

seems to be the best candidate, although high correlations are only found for FFL's with 'AND' gate interactions instead of feedback circuits. When performing a principal component analysis and looking at the distances between ΔT and the different topological features, the pattern we observe is similar to the previous one: s_0 lies close to ΔT for FFL circuits with 'AND' gates. Therefore, although we have not been able to find a general dynamical feature determining the amplitude of propagated oscillations, we still see that, given high correlations of the propagated noise with s_0 and ΔT, then high correlations are also found between ΔT and s_0.

4. Final discussion

To conclude this thesis we now summarize the main results it contains, while we try to emphasize the most relevant ideas. Although the reader has possibly found a clear distinction between the approaches taken to study the machinery involved in generating calcium oscillations, and the ones taken to characterize the response and dynamical features of simple gene circuits, there is an underlying approach which is shared by both studies. As stated in the introductory chapter, Biology has historically taken a reductionist approach, by trying to understand the behavior of biological entities through understanding that of the little components that form them. On the contrary, Systems Biology takes a different approach, in which the biological system is taken as a whole, with its properties emerging from the complex interactions between simple components. It is this 'systems' approach what is shared by both studies: we unveiled the interactions between the components forming the machinery responsible for the calcium oscillations by taking the system as a 'black box' and comparing its input-output response with that of diverse mathematical models describing it. Similarly, we studied the properties of simple gene networks by characterizing the input-output relationship of their responses, with mathematical modeling being applied later to explain the obtained results. Apart from the fact that both signal transduction and gene expression are related to the ability of cells to process information, perhaps this 'black box' approach represents the main common framework in which both studies take place.

Another less fundamental although still important approach which is common to both studies, is perhaps the fact that the extracted conclusions do not depend on the specific parameters the systems are operating in. Instead of fitting models to specific responses, we have taken a statistical approach in which responses have been analyzed in terms of their statistical behavior under large sets of random conditions. This allowed us to deal with the observed cell-to-cell variation in the microfluidic experiments, and made the conclusions extracted in the study of gene networks more general.

Once summarized the common points found along the whole thesis, we can go on and describe the specific results obtained in each section. To study the IP_3-mediated calcium signaling pathway we have followed a 'cellular interrogation' protocol, in which we asked some questions to real cells and compared their answers with the ones provided by mathematical models. In this way we can reject the models which do not give the correct answers and keep the others, with the hope of getting to a point in which only one model remains. The abil-

ity of this approach to ultimately get a unique answer depends on various factors, like the number of models and their types. Ideally, the more questions we pose the more precise we would be in our model selection, but this is limited by the number of models we have, as well as by their nature. In fact, when we put into practice this procedure in our calcium context, we got to a point in which two models remained, and we could not distinguish them with feasible experiments. Nevertheless, in this case both models describe a very similar structure of the network interactions, and therefore they both provide consistent information. To further test our results, we decided to reproduce the experimental trajectories with the selected models. When doing so we found that, although they were able to reproduce the typical calcium spiking pattern for both constant and pulsed stimulation, in the latter case they failed to do so when input pulses had relatively short periods, exhibiting spikes in between input pulses. This was not consistent with what we observed in real cells, and seemed to be caused by slow IP_3 decay in the models. When implementing fast IP_3 dynamics in the models, they were again able to follow fast pulses, suggesting that the assumption of fast IP_3 dynamics was correct. This was ultimately tested by performing a specific experiment that also gave us confidence about the protocol as a whole. The fact that the whole protocol follows a 'black box' strategy makes it flexible to be easily applied to study other cellular systems in which oscillations are naturally present, opening the door to use microfluidic devices and reverse engineering techniques for the study of such nonlinear systems.

To study the influence that the presence of specific interactions has on the behavior of a gene network, we have taken a simple three component module as a simple yet complete model. It represents a complete platform which is able to model different gene structures commonly found in real networks (like feedbacks or feedforwards), while its simplicity allows us to derive theoretical expressions which make good predictions of its behavior and provides with a deeper insight in the influence that each interaction has on the overall response of the network.

In the first part of this section we have analyzed the propagation of amplitude modulated and frequency modulated signals along the network, while taking into account how random fluctuations distort the information embedded in the input signal. When considering propagation of noiseless signals, we found that feedback circuits exhibit trade-offs between their ability to propagate both types of signals, while feedforward structures are able to overcome this limit, improving the propagation of amplitude and frequency modulated signals with respect to a simple linear cascade. By including noise in the analysis, we focused on the ability of the circuits to improve the signal to noise ratio or to filter noise. In this case we observed that a positive feedback interaction added to a linear cascade is able to increase the signal to noise ratio of amplitude signals, while a negative feedback tends to reduce that quantity. At the same time, coherent feedforward loops are able to improve signal to noise ratio for both amplitude and frequency modulated signals, while incoherent feedforward loops are

able to operate as high-pass filters, performing noise-filtering of oscillatory signals. A similar behavior was found for negative feedbacks, which are able to operate as low-pass filters, a property which helps them filtering high frequency noise while propagating slower signals. To conclude this study, we checked that the range of interaction strengths in which these circuits are able to perform the described tasks can be found in real systems by computing those strengths for models of real feedforward networks fitted to experimental data.

In the second part of this study we focused on understanding the connections between the topology of gene networks, their ability to propagate noise and oscillatory signals, and their response to transient changes in the input stimulus. To account for all these properties, we defined different quantities: the topology of the network was given by the strength of the interactions (susceptibilities and certain combinations of them). The ability to propagate frequency modulated signals was quantified by the bandwidth of the propagated oscillations, while propagated noise was given by the coefficient of variation of the output. Finally, the dynamic response to transient stimuli was given by the response times. When we examined the relationship between topology, propagation of oscillatory signals and response times, we first observed a strong correlation between the bandwidth of the propagated oscillations, the sum of response times and the feedback or feedforward strength. This last quantity had been previously stated to uniquely determine the range of frequencies propagated by the network, and these results confirmed the expected behavior. At the same time, a principal component analysis allowed us to find a clear connection between the sum of response times and the feedback or feedforward strength, thus closing the circle of the relationship between structure, dynamics, and signal propagation, when the latter refers to the ability to propagate oscillations. The same analysis was performed when the signal propagation feature was the amplitude of the noise propagated from the input to the output. In this case, correlations are not that clear: the overall susceptibility is strongly involved in determining the amplitude of the propagated fluctuations, but no clear dynamical feature seems to be related with this feature of the propagated signal for every circuit topology. In this sense, the difference between response times seems to be the best candidate, but high correlations are found only for feedforward circuits operating under 'AND' logic. Nevertheless, with the principal component analysis we found that circuits in which the amplitude of the fluctuations was strongly correlated with the response times difference, also exhibited a close connection between this dynamical feature and the overall susceptibility, thus, closing the circle again between structure, dynamics and signal propagation.

5. Discusión final

Como conclusión a esta tesis, exponemos a continuación un breve resumen de los resultados obtenidos, intentando resaltar las ideas más relevantes. El lector probablemente haya encontrado una clara distinción entre los métodos utilizados en el estudio de la vía de señalización de calcio intracelular y en la caracterización de la respuesta de circuitos genéticos simples a distintas señales. No obstante, ambos estudios comparten una filosofía común. Como se indicó en la introducción de la tesis, la Biología ha tendido a ser, históricamente, una disciplina reduccionista, al intentar explicar el comportamiento de diferentes entidades biológicas a base de entender cómo se comportan los elementos que las componen. Por el contrario, la Biología de Sistemas toma un enfoque diferente, en el que el sistema biológico estudiado es analizado como un todo cuyas propiedades emergen debido a las interacciones entre componentes más simples. Es precisamente esta aproximación lo que comparten ambos estudios: para descifrar la maquinaria celular que da lugar a las oscilaciones de calcio intracelular, tratamos el sistema como una "caja negra" y comparamos su comportamiento estímulo-respuesta con el de varios modelos matemáticos que describen el sistema. Del mismo modo, estudiamos las propiedades de circuitos genéticos simples caracterizando su comportamiento estímulo-respuesta, con el apoyo de modelos matemáticos que son utilizados a posteriori para entender los resultados obtenidos. Aparte del hecho de que ambos procesos, expresión genética y transducción de señal, ocurren en el mismo contexto dentro de la célula, quizá sea esta aproximación tipo "caja negra" lo que da un enfoque común a ambos estudios.

Otro enfoque no tan fundamental pero igualmente importante que se da en ambos estudios se refleja en el hecho de que todos los resultados obtenidos son independientes de los parámetros específicos en los que los sistemas puedan estar operando. En vez de ajustar los modelos a respuestas específicas, hemos llevado a cabo un estudio estadístico en el que las respuestas han sido analizadas en términos de su comportamiento estadístico para un amplio rango de parámetros aleatorios. Esto nos ha permitido, por un lado, evitar el problema de la variabilidad entre células en los experimentos hechos en dispositivos microfluídicos, y, por otro, obtener conclusiones más generales sobre el comportamiento de los circuitos genéticos.

Una vez expuesto el enfoque general llevado a cabo a lo largo de toda la tesis, resumimos ahora los resultados obtenidos en cada sección. Para estudiar la vía de señalización

de calcio mediado por IP_3 hemos llevado a cabo una especie de "interrogatorio celular", en el que diversas preguntas han sido planteadas a células vivas y sus respuestas han sido comparadas con las dadas por diferentes modelos matemáticos que pretenden describir este sistema en concreto. De este modo podemos descartar los modelos que no nos dan las respuestas adecuadas y quedarnos con los restantes, con la esperanza de llegar a un punto en el que lleguemos a tener sólo un modelo correcto. La capacidad de este método para llegar a obtener una respuesta única al problema depende de varios factores: idealmente, cuantas más preguntas hagamos al sistema más precisa será nuestra selección, aunque esto se verá limitado por el número de modelos disponibles y por su naturaleza. De hecho, cuando ponemos este método en práctica en el contexto del calcio intracelular, llegamos a un punto en el que tenemos dos modelos que no somos capaces de distinguir con ningún otro experimento. No obstante, en este caso ambos modelos dan una descripción muy similar de la estructura del sistema, por lo que la información proporcionada por ambos es consistente. Para comprobar los resultados obtenidos, hemos intentado reproducir las típicas trayectorias experimentales con nuestros modelos. Al hacer esto, hemos visto que, aunque en un principio ambos modelos son capaces de reproducir el comportamiento de las células cuando son sometidas a una concentración constante o a pulsos de histamina, en este último caso los modelos empiezan a fallar cuando el periodo de los pulsos se hace pequeño, dando lugar a la aparición de picos de calcio entre pulsos. Esto nunca lo vemos en experimentos reales, y parece ser debido a una degradación lenta del IP_3 en los modelos. Cuando modificamos estos para que el IP_3 tenga una dinámica rápida, los picos de calcio son capaces de ir en fase con los pulsos, recuperando así lo observado en los experimentos, y dándonos una buan confirmación de que la dinámica del IP_3 es efectivamente rápida. Esto ha sido comprobado con otro tipo de experimento diseñado expresamente para ello, lo que nos ha dado mayor confianza en el protocolo en general. El hecho de que todo el método se base en una estrategia de "caja negra" hace que sea flexible, y que su aplicación a otros sistemas oscilatorios sea relativamente sencialla, abriendo así la puerta al uso de dispositivos microfluídicos y a la aplicación de técnicas de ingeniería inversa al estudio de este tipo de sistemas no lineales.

Para estudiar la influencia de diferentes interacciones en el comportamiento de redes de genes, hemos elegido un circuito simple de tres genes como modelo. este modelo es completo en cuanto a que es capaz de modelizar diferentes estructures habituales en redes reales (feedbacks o feedforwards), mientras que su sencillez permite desarrollar expresiones teóricas que dan buenas predicciones del comportamiento del circuito, y que permiten entender más en profundidad la influencia que en él tiene cada una de las interacciones.

En la primera parte de esta sección hemos analizado la propagación de señales moduladas en amplitud y en frecuencia, a la vez que hemos tenido en cuenta la distorsión de la señal causada por fluctuaciones aleatorias en la concentración de los diferentes componentes.

Cuando consideramos la propagación de señales sin ruido, vemos que los feedbacks tienen un trade-off entre su capacidad de transmitir un tipo u otro de señal, mientras que los feedforwards son capaces de superar esa limitación y mejorar la transmisión de ambos tipos de señales respecto a una cascada lineal. Al incluir el ruido en el análisis, nos centramos en la capacidad de los circuitos para mejorar la relación señal/ruido o para filtrar fluctuaciones. En ete caso, hemos visto que añadiendo un feedback positivo a una cascada lineal aumenta la relación señal/ruido de señales propagadas en amplitud, mientras que un feedback negativo tiende a reducir esta relación. Al mismo tiempo, los feedforward loops coherentes son capaces de aumentar la relación señal/ruido, tanto para señales modulas en amplitud como en frecuencia, mientras que los feedforward loops incoherentes son capaces de funcionar como filtros de pasa-alto, lo que les permite filtrar el ruido de señales oscilatorias. Un comportamiento similar se da en los feedback negativos, que son capaces de funcionar como filtros de pasa-baja, filtrando fluctuaciones rápidas mientras permiten la propagación de oscilaciones más lentas. Este estudio lo hemos concluido comprobando en modelos ajustados a datos experimentales que, efectivamente, las redes de genes reales son capaces de operar en regímenes en los que las propiedades antes expuestas son susceptibles de encontrarse.

En la segunda parte de este estudio nos hemos centrado en la relación entre la topología de las redes genéticas, su capacidad para propagar ruido y señales oscilatorias, y su respuesta a cambios bruscos en la concentración de señal. Para cuantificar estas propiedades hemos definido varias magnitudes: la topología de la red viene dada por la fuerza de sus interacciones (susceptibilidades y combinaciones de éstas). La capacidad de propagar oscilaciones es caracterizada por la anchura de banda de las oscilaciones de la respuesta, mientras que el ruido propagado viene dado por el el coeficiente de variación de la respuesta. Por último, la respuesta dinámica a cambios en la señal ha sido medida por los tiempos de respuesta. Al examinar las conexiones entre la topología de los circuitos, su capacidad para propagar señales oscilatorias y sus tiempos de respuesta, lo primero que observamos es un alta correlación del ancho da banda de las oscilaciones propagadas con la suma de tiempos de respuesta y con la feedback o feedforward strength. Anteriormente ya habíamos mencionado que esta magnitud determinaba unívocamente la bandwidth de las oscilaciones, por lo que este resultado, en el que se observan siempre correlaciones 1, confirma una vez más esta afirmación. Al mismo tiempo, el análisis de componentes principales nos permite ver una clara relación entre la suma de tiempos de respuesta y la feedback o feedforward strength, por lo que de este modo se cierra el círculo de las relaciones entre estructura, dinámica y propagación de señal, al menos en el caso en el que esta última sea el ancho de banda de las oscilaciones de la respuesta. El mismo análisis se ha llevado a cabo cuando estudiamos la amplitud del ruido propagado como característica de la transmisión de señal. En este caso las correlaciones no son tan claras: la susceptibilidad total está profundamente relacionada con la amplitud de las fluctuaciones propagadas, pero esta característica de la propagación de señal no parece estar especialmente relacionada con ninguna propiedad dinámica del sistema. En este sentido, la

diferencia entre tiempos de respuesta parece ser la mejor candidata, pero sólo encontramos correlaciones altas para feedforwards con puertas 'AND'. No obstante, el análisis de componentes principales nos permite ver que aquellos circuitos en los que al amplitud de la fluctuaciones está altamente correlacionada con la diferencia de tiempos de respuesta, son precisamente aquellos que muestran una clara conexión entre esta característica dinámica y la susceptibilidad total. Por tanto, el círculo se cierra una vez más, mostrando la correlación entre estructura, dinámica y propagación de señal.

A. Calcium models and figures:

A.1. Equations for the calcium models

From previously published material, eight different models are tested. Among the models under study there are old models like Goldbeter [89] and Meyer [88] and newer ones like the 2 and 3 state Sneyd-LeBeau model [92]. Atri [90] and Li-Rinzel [91] models are tested in their class 1 and class 2 versions, as presented by Sneyd [68].

Although the specific interactions differ from one model to another, the basic species and cellular compartments interacting to give rise to calcium oscillations are the same among all of them. Therefore, endoplasmic reticulum (ER) calcium concentration, inositol trisphosphate (IP_3), calcium pumps or IP_3 receptors in the ER are common ingredients on most of the models. Schematic representations of each model are given in Fig. (2.15).

species	variable	units
cystosolic calcium	cc	μM
ER calcium	cer	μM
IP_3	p	μM
fraction of active IP_3 receptors	n	$u.l.$
fraction of receptors in state x	x	$u.l.$
fraction of receptors in state y	y	$u.l.$
fraction of receptors in state y_2	$y_2 = x + y$	$u.l.$

Table A.1.: Species on the different models and their associated variable

A.1.1. Meyer

$$\begin{aligned} \frac{dp}{dt} &= k_+(cc) - k_-(p) \\ \frac{dcer}{dt} &= J_2(cer, cc) - J_1(p, cer) \\ \frac{dcc}{dt} &= J_1(p, cer) - J_2(cer, cc) - mitochondria(cc) \end{aligned} \tag{A.1}$$

$$\begin{aligned} k_+(cc) &= \frac{c_4 R}{cc + k_3} \cdot cc \\ k_-(p) &= c_5 \cdot p \\ J_1(p, cer) &= \frac{c_1 \cdot cer \cdot p^3}{(k_1 + p)^3} \\ J_2(cer, cc) &= \frac{c_2 \cdot cc^2}{(cc + k_2)^2} - c_3 \cdot cer^2 \\ mitochondria(cc) &= c_6 (\frac{cc}{c_7})^{3.3} - c_6 \end{aligned} \tag{A.2}$$

parameter	original value
c_1	6.64 s^{-1}
k_1	0.1 μM
c_2	5 $\mu M s^{-1}$
k_2	0.15 μM
c_3	3.13·10^{-5} $(\mu M s)^{-1}$
c_4	1 s^{-1}
k_3	1 μM
c_5	2 s^{-1}
c_6	0.5 $\mu M s^{-1}$
c_7	0.6 μM
R^*	0.31 μM

Table A.2.: Meyer: original parameters. * indicates the input parameter.

A.1.2. Goldbeter

$$\begin{aligned} \frac{dcer}{dt} &= v_2(cc) - v_3(cc, cer) - k_f \cdot cer \\ \frac{dcc}{dt} &= v_0 + v_1 \cdot b - v_2(cc) + v_3(cc, cer) + k_f \cdot cer - k \cdot cc \end{aligned} \tag{A.3}$$

$$\begin{aligned} v_2(cc) &= V_{M2} \cdot \frac{cc^2}{k_2^2 + cc^2} \\ v_3(cc, cer) &= V_{M3} \cdot \frac{cer^2}{K_R^2 + cer^2} \cdot \frac{cc^4}{K_A^4 + cc^4} \end{aligned} \tag{A.4}$$

parameter	original value
v_0	1 $\mu M s^{-1}$
k	10 s^{-1}
k_f	1 s^{-1}
v_1^*	7.3 $\mu M s^{-1}$
V_{M1}	65 $\mu M s^{-1}$
V_{M2}	500 $\mu M s^{-1}$
k_2	1 μM
K_R	2 μM
K_A	0.9 μM
b	0.31 $u.l.$

Table A.3.: Goldbeter: original parameters. * indicates the input parameter.

A.1.3. Atri1

$$\begin{aligned} \frac{dn}{dt} &= \frac{1}{\tau_n} \cdot (n_{inf}(cc) - n) \\ \frac{dp}{dt} &= ir \cdot (p_{st} - p) \\ \frac{dcer}{dt} &= J_{cc} - J_{IP3} \\ \frac{dcc}{dt} &= J_{IP3} - J_{cc} + \delta(J_{in} - J_{out}) \end{aligned} \tag{A.5}$$

$$\begin{aligned} J_{cc} &= \frac{\gamma_2 \cdot cc}{k_{\gamma_2} + cc} \\ J_{IP3} &= k_{flux}\mu(p) \cdot n \left(b + \frac{V_1 cc}{k_1 + cc}\right)(\gamma \cdot cer - cc) \\ J_{in} &= \beta_1 + \beta_2 \cdot p_{st} \\ J_{out} &= \frac{\gamma_1 cc^2}{k_{\gamma_1}^2 + cc^2} \\ n_{inf}(cc) &= 1 - \frac{cc^2}{k_2^2 + cc^2} \\ \mu(p) &= \mu_0 + \frac{\mu_1 \cdot p}{k_\mu + p} \end{aligned} \tag{A.6}$$

parameter	original value
p^*_{st}	10 μM
β_1	1 $\mu M s^{-1}$
β_2	0.2 s^{-1}
ir	0.08 s^{-1}
k_{flux}	4.8 $\mu M s^{-1}$
δ	0.01 $u.l.$
γ_1	24 $\mu M s^{-1}$
γ_2	20 $\mu M s^{-1}$
k_{γ_1}	0.4 μM
k_{γ_2}	0.06 μM
γ	5.405 μM^{-1}
τ_n	2 s
k_2	0.7 μM
b	0.111 $u.l.$
V_1	0.889 $u.l.$
k_1	0.7 μM
k_μ	4 μM
μ_0	0.567 $u.l.$
μ_1	0.433 $u.l.$

Table A.4.: Atri1: original parameters. * indicates the input parameter.

A.1.4. Atri2

$$\begin{aligned} \frac{dp}{dt} &= v_4 \left[\frac{cc + (1-\alpha)k_4}{cc + k_4} \right] - ir \cdot p \\ \frac{dcer}{dt} &= J_{cc} - J_{IP3} \\ \frac{dcc}{dt} &= J_{IP3} - J_{cc} + \delta(J_{in} - J_{out}) \end{aligned} \tag{A.7}$$

$$
\begin{aligned}
J_{cc} &= \frac{\gamma_2 \cdot cc}{k_{\gamma_2} + cc} \\
J_{IP3} &= k_{flux}\mu(p) \cdot n_{inf}(cc) \left(b + \frac{V_1 cc}{k_1 + cc} \right) (\gamma \cdot cer - cc) \\
J_{in} &= \beta_1 + \beta_2 \cdot p_{st} \\
J_{out} &= \frac{\gamma_1 cc^2}{k_{\gamma_1}^2 + cc^2} \\
n_{inf}(cc) &= 1 - \frac{cc^2}{k_2^2 + cc^2} \\
\mu(p) &= \mu_0 + \frac{\mu_1 \cdot p}{k_\mu + p}
\end{aligned}
\tag{A.8}
$$

parameter	original value
v_4^*	6 $\mu M s^{-1}$
β_1	1 $\mu M s^{-1}$
β_2	0.2 s^{-1}
k_{flux}	4.8 $\mu M s^{-1}$
α	0.97 $u.l.$
δ	0.01 $u.l.$
γ_1	24 $\mu M s^{-1}$
γ_2	20 $\mu M s^{-1}$
k_{γ_1}	0.4 μM
k_{γ_2}	0.06 μM
γ	5.405 μM^{-1}
τ_n	2 s
k_2	0.7 μM
b	0.111 $u.l.$
V_1	0.889 $u.l.$
k_1	0.7 μM
k_μ	4 μM
μ_0	0.567 $u.l.$
μ_1	0.433 $u.l.$
k_4	1.1 μM
ir	0.08 s^{-1}

Table A.5.: Atri2: original parameters. * indicates the input parameter.

A.1.5. Li-Rinzel1

$$\begin{aligned}
\frac{dn}{dt} &= A(K_d - (cc + K_d) \cdot n) \\
\frac{dp}{dt} &= ir \cdot (p_{st} - p) \\
\frac{dcer}{dt} &= J_{cc} - J_{IP3} - J_{noIP3} \\
\frac{dcc}{dt} &= J_{IP3} + J_{noIP3} - J_{cc} + eps \cdot (J_{in} - J_{out})
\end{aligned} \tag{A.9}$$

$$\begin{aligned}
J_{cc} &= \frac{V_e cc^2}{K_e^2 + cc^2} \\
J_{IP3} &= P \left(\frac{p \cdot n \cdot cc}{(p + K_i) \cdot (cc + K_a)} \right)^3 \cdot (\sigma^{-1} cer - cc) \\
J_{noIP3} &= L(\sigma^{-1} cer - cc) \\
J_{in} &= \alpha_1 + \alpha_2 \cdot p_{st} \\
J_{out} &= \frac{V_p cc^2}{K_p^2 + cc^2}
\end{aligned} \tag{A.10}$$

parameter	original value
p_{st}^*	0.8 μM
α_1	1 $\mu M s^{-1}$
α_2	0.25 s^{-1}
eps	0.01 $u.l.$
ir	0.02 s^{-1}
L	$9.25 \cdot 10^{-3}$ s^{-1}
P	66.6 s^{-1}
K_i	1 μM
K_a	0.4 μM
V_e	1 $\mu M s^{-1}$
K_e	0.2 μM
A	0.5 s^{-1}
K_d	0.4 μM
σ	0.185 $u.l.$
V_p	5 $\mu M s^{-1}$
K_p	0.3 μM

Table A.6.: Li-Rinzel1: original parameters. * indicates the input parameter.

A.1.6. Li-Rinzel2

$$\begin{aligned} \frac{dp}{dt} &= v_4 \left[\frac{cc + (1-\alpha)k_4}{cc + k_4} \right] - ir \cdot p \\ \frac{dcer}{dt} &= J_{cc} - J_{IP3} - J_{noIP3} \\ \frac{dcc}{dt} &= J_{IP3} + J_{noIP3} - J_{cc} + eps \cdot (J_{in} - J_{out}) \end{aligned} \tag{A.11}$$

$$\begin{aligned} J_{cc} &= \frac{V_e cc^2}{K_e^2 + cc^2} \\ J_{IP3} &= P \left(\frac{p \cdot n(cc) \cdot cc}{(p + K_i) \cdot (cc + K_a)} \right)^3 \cdot (\sigma^{-1} cer - cc)) \\ J_{noIP3} &= L(\sigma^{-1} cer - cc) \\ J_{in} &= \alpha_1 + \alpha_2 \cdot p_{st} \\ J_{out} &= \frac{V_p cc^2}{K_p^2 + cc^2} \\ n(cc) &= \frac{K_d}{cc + K_d} \end{aligned} \tag{A.12}$$

parameter	original value
v_4^*	0.7 $\mu M s^{-1}$
α	0.97 $u.l.$
k_4	1.1 μM
ir	0.02 s^{-1}
α_1	1 $\mu M s^{-1}$
α_2	0.25 s^{-1}
eps	0.01 $u.l.$
L	$9.25 \cdot 10^{-3}$ s^{-1}
P	66.6 s^{-1}
K_i	1 μM
K_a	0.4 μM
V_e	1 $\mu M s^{-1}$
K_e	0.2 μM
A	0.5 s^{-1}
K_d	0.4 μM
σ	0.185 $u.l.$
V_p	5 $\mu M s^{-1}$
K_p	0.3 μM

Table A.7.: Li-Rinzel2: original parameters. * indicates the input parameter.

A.1.7. Sneyd-LeBeau

3 state:

$$\begin{aligned} \frac{dx}{dt} &= \phi_{-1}(cc) \cdot y - P\phi_1(cc) \cdot x + \phi_3(cc)(1 - x - y) \\ \frac{dy}{dt} &= P\phi_1(cc) \cdot x - \phi_{-1}(cc) \cdot y - \phi_2(cc)y \\ \frac{dcc}{dt} &= k_f y^4 - \frac{V_p cc^2}{K_p^2 + cc^2} + J_{leak} \end{aligned} \tag{A.13}$$

2 state:

$$\begin{aligned} \frac{dy_2}{dt} &= \phi_3(cc)(1 - y_2) - \left(\frac{\phi_1(cc)\phi_2(cc)P}{\phi_1(cc)P + \phi_{-1}(cc)} \right) \cdot y_2 \\ \frac{dcc}{dt} &= k_f \left(\frac{Py_2\phi_1(cc)}{\phi_1(cc)P + \phi_{-1}(cc)} \right)^4 - \frac{V_p cc^2}{K_p^2 + cc^2} + J_{leak} \end{aligned} \tag{A.14}$$

$$\begin{aligned} \phi_1(cc) &= \frac{(k_1 R_1 + r_2 cc)}{R_1 + cc} \\ \phi_{-1}(cc) &= \frac{(k_{-1} + r_{-2}) R_3}{R_3 + cc} \\ \phi_2(cc) &= \frac{(k_2 R_3 + r_4 cc)}{R_3 + cc} \\ \phi_3(cc) &= \frac{(k_3 R_5 + r_6 cc)}{R_5 + cc} \end{aligned} \tag{A.15}$$

parameter	original value
R_1	6 μM
r_2	100 s^{-1}
R_3	50 μM
r_4	20 s^{-1}
R_5	1.6 μM
r_6	0 s^{-1}
r_{-2}	0 s^{-1}
k_1	0 s^{-1}
k_2	0.53 s^{-1}
k_3	1 s^{-1}
k_{-1}	0.88 s^{-1}
k_f	28 $\mu M s^{-1}$
V_p	1.2 $\mu M s^{-1}$
J_{leak}	0.2 $\mu M s^{-1}$
K_p	0.18 μM
P^*	0.47 $u.l.$

Table A.8.: Sneyd-LeBeau: original parameters. * indicates the input parameter.

A.2. Workflow of the calcium model selection

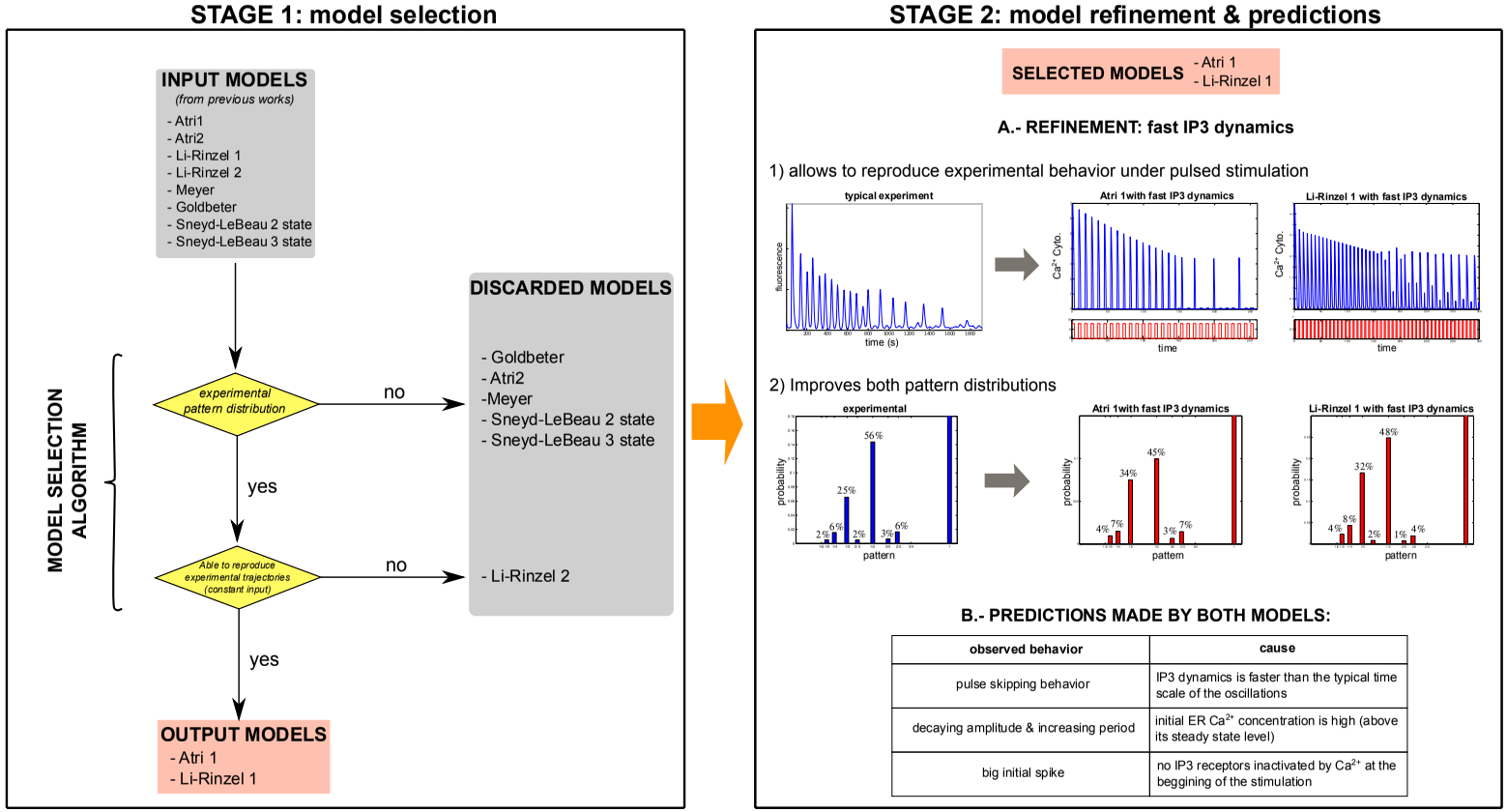


Figure A.1.: Overall process: flux diagram showing the overall model selection process.

B. Gene circuits linear approximations:

B.1. Dynamic evolution of simple 3 component circuits

B.1.1. Two component interactions

Linear approximation for any 2 component interactions.

$$\frac{d\Delta \boldsymbol{x}}{dt} = \boldsymbol{M} \cdot \Delta \boldsymbol{x} \tag{B.1}$$

$$\boldsymbol{M} = -\begin{pmatrix} \frac{H_{11}}{\tau_1} & \frac{H_{12}}{\tau_1} \\ \frac{H_{21}}{\tau_2} & \frac{H_{22}}{\tau_2} \end{pmatrix} \tag{B.2}$$

$$\Delta \boldsymbol{x} = \begin{pmatrix} \delta x_1 \\ \delta x_2 \end{pmatrix} \tag{B.3}$$

Different time scales

$\tau_1 \neq \tau_2$

$$\begin{aligned} \delta x_2(t) \quad = \quad & \frac{\delta x_2(0)}{\Delta} e^{-\frac{1}{2}(\frac{H_{11}}{\tau_1}+\frac{H_{22}}{\tau_2})t}[\Delta cosh(\frac{t\Delta}{2}) + (\frac{H_{11}}{\tau_1} - \frac{H_{22}}{\tau_2})sinh(\frac{t\Delta}{2})] - \\ & -2\frac{\delta x_1(0)}{\Delta} e^{-\frac{1}{2}(\frac{H_{11}}{\tau_1}+\frac{H_{22}}{\tau_2})t} \frac{H_{21}}{\tau_1} senh(\frac{t\Delta}{2}) \end{aligned} \tag{B.4}$$

With:

$$\Delta = \sqrt{Tr^2 - 4Jac} \tag{B.5}$$

$$Tr = -(\frac{H_{11}}{\tau_1} + \frac{H_{22}}{\tau_2}) \tag{B.6}$$

Same timescales

$\tau_1 = \tau_2 = \tau = \frac{1}{\delta}$

$$\delta x_2(t) = e^{-\delta t}[\delta x_2(0) cosh(\sqrt{\frac{H_{21}H_{12}}{\tau^2}}t) + \delta x_1(0)\sqrt{\frac{H_{21}}{H_{12}}} sinh(\sqrt{\frac{H_{21}H_{12}}{\tau^2}}t)] \tag{B.7}$$

B.1.2. Linear Cascade

$$\boldsymbol{M} = -\begin{pmatrix} \frac{H_{00}}{\tau_0} & 0 & 0 \\ \frac{H_{10}}{\tau_1} & \frac{H_{11}}{\tau_1} & 0 \\ 0 & \frac{H_{21}}{\tau_2} & \frac{H_{22}}{\tau_2} \end{pmatrix} \tag{B.8}$$

No auto-regulation implies: $H_{00} = H_{11} = H_{22}$
Total susceptibility: $S_o = H_{10}H_{21}$

Different time scales

$\delta_0 \neq \delta_1 \neq \delta_2$

$$\begin{aligned} \delta y(t) \quad = \quad & \delta y(0)e^{-\delta_2 t} + \delta x(0) H_{21}\delta_2 \frac{e^{-\delta_1 t} - e^{-\delta_2 t}}{\delta_1 - \delta_2} + \\ & + \delta s(0) H_{10}H_{21}\delta_1\delta_2 (\frac{e^{-\delta_0 t}}{(\delta_0 - \delta_1)(\delta_0 - \delta_2)} + \\ & + \frac{e^{-\delta_1 t}}{(\delta_1 - \delta_0)(\delta_1 - \delta_2)} + \frac{e^{-\delta_2 t}}{(\delta_2 - \delta_0)(\delta_2 - \delta_1)}) \end{aligned} \tag{B.9}$$

Same time scales

$\delta_0 = \delta_1 = \delta_2 = \delta$

$$\delta y(t) = \delta y(0)e^{-\delta t} - \delta x(0) H_{21}\delta t e^{-\delta t} + \delta s(0)\frac{H_{10}H_{21}}{2} t^2\delta^2 e^{-\delta t} \tag{B.10}$$

B.1.3. Feedbacks

$$\boldsymbol{M} = -\begin{pmatrix} \frac{H_{00}}{\tau_0} & 0 & 0 \\ \frac{H_{10}}{\tau_1} & \frac{H_{11}}{\tau_1} & \frac{H_{12}}{\tau_1} \\ 0 & \frac{H_{21}}{\tau_2} & \frac{H_{22}}{\tau_2} \end{pmatrix} \tag{B.11}$$

Again, no auto-regulation implies: $H_{00} = H_{11} = H_{22} = 1$

The only easy solutions are found when all time scales are the same, thus, given $\delta_0 = \delta_1 = \delta_2 = \delta$:

Positive FB mutual activation ($s_{12} > 0$ and $s_{21} > 0$)

$$\begin{aligned}\delta y(t) = e^{-\delta t} \quad & [\delta y(0) \cdot cosh(\sqrt{s_{21}s_{12}}\delta t) + \\ & +\delta x(0)\sqrt{\frac{s_{21}}{s_{12}}} \cdot sinh(\sqrt{s_{21}s_{12}}\delta t) + \\ & +\delta s(0)\frac{s_{21}}{s_{12}} \cdot [cosh(\sqrt{s_{21}s_{12}}\delta t) - 1]]\end{aligned} \tag{B.12}$$

Positive FB mutual inhibition ($s_{12} < 0$ and $s_{21} < 0$)

$$\begin{aligned}\delta y(t) = e^{-\delta t} \quad & [\delta y(0) \cdot cosh(\sqrt{s_{21}s_{12}}\delta t) - \\ & -\delta x(0)\sqrt{\frac{s_{21}}{s_{12}}} \cdot sinh(\sqrt{s_{21}s_{12}}\delta t) - \\ & -\delta s(0)\frac{s_{21}}{s_{12}} \cdot [cosh(\sqrt{s_{21}s_{12}}\delta t) - 1]]\end{aligned} \tag{B.13}$$

Negative FB ($s_{12} > 0$ and $s_{21} < 0$)

$$\begin{aligned}\delta y(t) = e^{-\delta t} \quad & [\delta y(0) \cdot cos(\sqrt{s_{21}s_{12}}\delta t) - \\ & -\delta x(0)\sqrt{\frac{s_{21}}{s_{12}}} \cdot sin(\sqrt{s_{21}s_{12}}\delta t) - \\ & -\delta s(0)\frac{s_{21}}{s_{12}} \cdot [cos(\sqrt{s_{21}s_{12}}\delta t) - 1]]\end{aligned} \tag{B.14}$$

Negative FB ($s_{12} < 0$ and $s_{21} > 0$)

$$\begin{aligned}\delta y(t) = e^{-\delta t} \quad & [\delta y(0) \cdot cos(\sqrt{s_{21}s_{12}}\delta t) + \\ & +\delta x(0)\sqrt{\frac{s_{21}}{s_{12}}} \cdot sin(\sqrt{s_{21}s_{12}}\delta t) - \\ & +\delta s(0)\frac{s_{21}}{s_{12}} \cdot [cos(\sqrt{s_{21}s_{12}}\delta t) - 1]]\end{aligned} \tag{B.15}$$

B.1.4. Feed-forward loops

$$\boldsymbol{M} = -\begin{pmatrix} \frac{H_{00}}{\tau_0} & 0 & 0 \\ \frac{H_{10}}{\tau_1} & \frac{H_{11}}{\tau_1} & 0 \\ \frac{H_{20}}{\tau_2} & \frac{H_{21}}{\tau_2} & \frac{H_{22}}{\tau_2} \end{pmatrix} \tag{B.16}$$

Different time scales

$\delta_0 \neq \delta_1 \neq \delta_2$

$$\begin{aligned}\delta y(t) &= \delta y(0)e^{-\delta_2 t} + \delta x(0)H_{21}\delta_2\frac{e^{-\delta_1 t} - e^{-\delta_2 t}}{\delta_1 - \delta_2} + \\ &+ \delta s(0)[e^{-\delta_0 t}\frac{H_{20}\delta_2}{\delta_0 - \delta_2} - e^{-\delta_2 t}\frac{H_{20}\delta_2}{\delta_0 - \delta_2} + \\ &\quad + e^{-\delta_0 t}\frac{H_{10}H_{21}\delta_1\delta_2}{(\delta_0 - \delta_1)(\delta_1 - \delta_2)} + \\ &\quad + e^{-\delta_1 t}\frac{H_{10}H_{21}\delta_1\delta_2}{(\delta_1 - \delta_0)(\delta_1 - \delta_2)} + e^{-\delta_2 t}\frac{H_{10}H_{21}\delta_1\delta_2}{(\delta_2 - \delta_0)(\delta_2 - \delta_1)}]\end{aligned} \tag{B.17}$$

Same time scales

$\delta_0 = \delta_1 = \delta_2 = \delta$

$$\delta y(t) = \delta y(0)e^{-\delta t} - \delta x(0)H_{21}\delta t e^{-\delta t} - \delta s(0)e^{-\delta t}(H_{20}\delta t - \frac{H_{10}H_{21}}{2}t^2\delta^2) \tag{B.18}$$

Bibliography

[1] J. Adler and W. W. Tso. "decision"-making in bacteria: chemotactic response of escherichia coli to conflicting stimuli. Science **184**, 1292 (1974). (cited on p. 1)

[2] S. Artavanis-Tsakonas, M. D. Rand, and R. J. Lake. Notch signaling: cell fate control and signal integration in development. Science **284**, 770 (1999). (cited on p. 1)

[3] R. S. Herbst. Review of epidermal growth factor receptor biology. Int J Radiat Oncol Biol Phys **59**, 21 (2004). (cited on p. 1)

[4] B. H. Natelson, J. Holaday, J. Meyerhoff, and P. E. Stokes. Temporal changes in growth hormone, cortisol, and glucose: relation to light onset and behavior. Am J Physiol **229**, 409 (1975). (cited on p. 1)

[5] G. M. C. . R. E. Hausman. *The Cell: a Molecular Approach* (2008). (cited on pages 1 and 66)

[6] D. Wylie, A. Stock, C. Y. Wong, and J. Stock. Sensory transduction in bacterial chemotaxis involves phosphotransfer between che proteins. Biochem Biophys Res Commun **151**, 891 (1988). (cited on p. 6)

[7] R. Belas, M. Simon, and M. Silverman. Regulation of lateral flagella gene transcription in vibrio parahaemolyticus. J Bacteriol **167**, 210 (1986). (cited on p. 6)

[8] S. M. K. S. P. P. Reed Larsen, Henry M. Kronenberg. *Williams Textbook of Endocrinology* (Saunders, 2002). (cited on pages 6 and 10)

[9] L. Wu and J. G. Belasco. Let me count the ways: mechanisms of gene regulation by mirnas and sirnas. Mol Cell **29**, 1 (2008). (cited on p. 6)

[10] T. Watanabe, Y. Totoki, A. Toyoda, M. Kaneda, S. Kuramochi-Miyagawa, Y. Obata, H. Chiba, Y. Kohara, T. Kono, T. Nakano, M. A. Surani, Y. Sakaki, and H. Sasaki. Endogenous sirnas from naturally formed dsrnas regulate transcripts in mouse oocytes. Nature **453**, 539 (2008). (cited on p. 6)

[11] E. J. Sontheimer and R. W. Carthew. Silence from within: endogenous sirnas and mirnas. Cell **122**, 9 (2005). (cited on p. 6)

[12] A. Ma'ayan, S. L. Jenkins, S. Neves, A. Hasseldine, E. Grace, B. Dubin-Thaler, N. J. Eungdamrong, G. Weng, P. T. Ram, J. J. Rice, A. Kershenbaum, G. A. Stolovitzky, R. D. Blitzer, and R. Iyengar. Formation of regulatory patterns during signal propagation in a mammalian cellular network. Science **309**, 1078 (2005). (cited on p. 7)

[13] S. H. Strogatz. Exploring complex networks. Nature **410**, 268 (2001). (cited on p. 7)

[14] Barabasi and Albert. Emergence of scaling in random networks. Science **286**, 509 (1999). (cited on p. 7)

[15] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A. L. Barabási. The large-scale organization of metabolic networks. Nature **407**, 651 (2000). (cited on p. 7)

[16] M. E. Newman. The structure of scientific collaboration networks. Proc Natl Acad Sci U S A **98**, 404 (2001). (cited on p. 7)

[17] A. Vázquez, R. Pastor-Satorras, and A. Vespignani. Large-scale topological and dynamical properties of the internet. Phys Rev E Stat Nonlin Soft Matter Phys **65**, 066130 (2002). (cited on p. 7)

[18] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: Simple building blocks of complex networks. Science **298**, 824 (2002). (cited on pages 7, 8, and 66)

[19] N. Rosenfeld, M. B. Elowitz, and U. Alon. Negative autoregulation speeds the response times of transcription networks. J. Mol. Biol. **323**, 785 (2002). (cited on pages 8 and 66)

[20] N. Rosenfeld and U. Alon. Response delays and the structure of transcription networks. J Mol Biol **329**, 645 (2003). (cited on pages 8, 16, and 66)

[21] A. Csikász-Nagy, O. Kapuy, A. Tóth, C. Pál, L. J. Jensen, F. Uhlmann, J. J. Tyson, and B. Novák. Cell cycle regulation by feed-forward loops coupling transcription and phosphorylation. Mol Syst Biol **5**, 236 (2009). (cited on pages 8, 9, and 66)

[22] S. Basu, Y. Gerchman, C. H. Collins, F. H. Arnold, and R. Weiss. A synthetic multicellular system for programmed pattern formation. Nature **434**, 1130 (2005). (cited on pages 8, 93, 94, and 105)

[23] K. E. Iles and H. J. Forman. Macrophage signaling and respiratory burst. Immunol Res **26**, 95 (2002). (cited on p. 9)

[24] D. A. Lauffenburger and J. J. Linderman. *Receptors: Models for Binding, Trafficking, and Signalling* (New York: Oxford University Press, 1993). (cited on pages 9, 24, and 25)

[25] N. Geva-Zatorsky, N. Rosenfeld, S. Itzkovitz, R. Milo, A. Sigal, E. Dekel, T. Yarnitzky, Y. Liron, P. Polak, G. Lahav, and U. Alon. Oscillations and variability in the p53 system. Mol Syst Biol **2**, 2006.0033 (2006). (cited on p. 9)

[26] J. E. Purvis, K. W. Karhohs, C. Mock, E. Batchelor, A. Loewer, and G. Lahav. p53 dynamics control cell fate. Science **336**, 1440 (2012). (cited on p. 9)

[27] S. Benloucif, M. J. Guico, K. J. Reid, L. F. Wolfe, M. L'hermite-Balériaux, and P. C. Zee. Stability of melatonin and temperature as circadian phase markers and their relation to sleep times in humans. J Biol Rhythms **20**, 178 (2005). (cited on p. 10)

[28] R. Guantes and J. F. Poyatos. Dynamical principles of two-component genetic oscillators. PLoS Comput Biol **2**, e30 (2006). (cited on p. 10)

[29] U. Alon, M. G. Surette, N. Barkai, and S. Leibler. Robustness in bacterial chemotaxis. Nature **397**, 168 (1999). (cited on p. 10)

[30] S. Arganda, R. Guantes, and G. G. de Polavieja. Sodium pumps adapt spike bursting to stimulus statistics. Nat Neurosci **10**, 1467 (2007). (cited on p. 10)

[31] L. Goentoro, O. Shoval, M. W. Kirschner, and U. Alon. The incoherent feedforward loop can provide fold-change detection in gene regulation. Mol Cell **36**, 894 (2009). (cited on p. 10)

[32] W. Ma, A. Trusina, H. El-Samad, W. A. Lim, and C. Tang. Defining network topologies that can achieve biochemical adaptation. Cell **138**, 760 (2009). (cited on p. 10)

[33] P. Cironi, I. A. Swinburne, and P. A. Silver. Enhancement of cell type specificity by quantitative modulation of a chimeric ligand. J Biol Chem **283**, 8469 (2008). (cited on p. 11)

[34] J. Paulsson. Models of stochastic gene expression. Physics of Life Reviews **2**, 157 (2005). (cited on pages 11, 18, 21, 69, 71, and 72)

[35] P. Nelson. *Biological Physics. Energy, Information, Life* (Freeman, 2004). (cited on p. 11)

[36] S. R. McGuffee and A. H. Elcock. Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. PLoS Comput Biol **6**, e1000694 (2010). (cited on p. 11)

[37] A. Novick and M. Weiner. Enzyme induction as an all-or-none phenomenon. Proc Natl Acad Sci U S A **43**, 553 (1957). (cited on p. 11)

[38] M. S. Ko, H. Nakauchi, and N. Takahashi. The dose dependence of glucocorticoid-inducible gene expression results from changes in the number of transcriptionally active templates. EMBO J **9**, 2835 (1990). (cited on p. 11)

[39] J. Paulsson and M. Ehrenberg. Random signal fluctuations can reduce random fluctuations in regulated components of chemical regulatory networks. Phys Rev Lett **84**, 5447 (2000). (cited on p. 11)

[40] J. Paulsson, O. G. Berg, and M. Ehrenberg. Stochastic focusing: fluctuation-enhanced sensitivity of intracellular regulation. Proc Natl Acad Sci U S A **97**, 7148 (2000). (cited on p. 11)

[41] M. B. Elowitz and S. Leibler. A synthetic oscillatory network of transcriptional regulators. Nature **403**, 335 (2000). (cited on p. 11)

[42] A. Becskei and L. Serrano. Engineering stability in gene networks by autoregulation. Nature **405**, 590 (2000). (cited on p. 11)

[43] H. H. McAdams and A. Arkin. Stochastic mechanisms in gene expression. Proc Natl Acad Sci U S A **94**, 814 (1997). (cited on pages 12 and 18)

[44] A. Arkin, J. Ross, and H. H. McAdams. Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected escherichia coli cells. Genetics **149**, 1633 (1998). (cited on p. 12)

[45] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. The Journal of Physical Chemistry **81**, 2340 (1977). (cited on pages 12, 20, and 73)

[46] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain. Stochastic gene expression in a single cell. Science **297**, 1183 (2002). (cited on p. 12)

[47] E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden. Regulation of noise in the expression of a single gene. Nat Genet **31**, 69 (2002). (cited on p. 12)

[48] A. Becskei, B. B. Kaufmann, and A. van Oudenaarden. Contributions of low molecule number and chromosomal positioning to stochastic gene expression. Nat Genet **37**, 937 (2005). (cited on p. 12)

[49] D. Volfson, J. Marciniak, W. J. Blake, N. Ostroff, L. S. Tsimring, and J. Hasty. Origins of extrinsic variability in eukaryotic gene expression. Nature **439**, 861 (2006). (cited on p. 12)

[50] W. J. Blake, G. Balázsi, M. A. Kohanski, F. J. Isaacs, K. F. Murphy, Y. Kuang, C. R. Cantor, D. R. Walt, and J. J. Collins. Phenotypic consequences of promoter-mediated transcriptional noise. Mol Cell **24**, 853 (2006). (cited on p. 12)

[51] C. Queitsch, T. A. Sangster, and S. Lindquist. Hsp90 as a capacitor of phenotypic variation. Nature **417**, 618 (2002). (cited on p. 12)

[52] D. A. Charlebois, N. Abdennur, and M. Kaern. Gene expression noise facilitates adaptation and drug resistance independently of mutation. Phys Rev Lett **107**, 218101 (2011). (cited on p. 12)

[53] S. F. Elena and R. E. Lenski. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nat Rev Genet **4**, 457 (2003). (cited on p. 12)

[54] M. Ptashne. On the use of the word 'epigenetic'. Curr Biol **17**, R233 (2007). (cited on p. 12)

[55] T. Egawa and D. R. Littman. Transcription factor ap4 modulates reversible and epigenetic silencing of the cd4 gene. Proc Natl Acad Sci U S A **108**, 14873 (2011). (cited on p. 13)

[56] W. C. Ratcliff, R. F. Denison, M. Borrello, and M. Travisano. Experimental evolution of multicellularity. Proc Natl Acad Sci U S A **109**, 1595 (2012). (cited on p. 13)

[57] A. J. Lotka. Contribution to the theory of periodic reactions. The Journal of Physical Chemistry **14**, 271 (1909). (cited on p. 13)

[58] A. L. Hodgkin and A. F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. 1952. Bull Math Biol **52**, 25 (1990). (cited on p. 13)

[59] L. Michaelis, M. L. Menten, K. A. Johnson, and R. S. Goody. The original michaelis constant: translation of the 1913 michaelis-menten paper. Biochemistry **50**, 8264 (2011). (cited on p. 13)

[60] J. Gunawardena. A linear framework for time-scale separation in nonlinear biochemical systems. PLoS One **7**, e36321 (2012). (cited on p. 14)

[61] G. E. Briggs and J. B. Haldane. A note on the kinetics of enzyme action. Biochem J **19**, 338 (1925). (cited on p. 14)

[62] M. Kaern, T. C. Elston, W. J. Blake, and J. J. Collins. Stochasticity in gene expression: from theories to phenotypes. Nat Rev Genet **6**, 451 (2005). (cited on pages 14 and 15)

[63] U. Alon. *An Introduction to Systems Biology: Design Principles of Biological Circuits* (CRCpress, 2006). (cited on pages 16 and 87)

[64] T. Höfer, L. Venance, and C. Giaume. Control and plasticity of intercellular calcium waves in astrocytes: a modeling approach. J Neurosci **22**, 4850 (2002). (cited on p. 16)

[65] D. J. Wilkinson. Stochastic modelling for quantitative description of heterogeneous biological systems. Nat Rev Genet **10**, 122 (2009). (cited on pages 16 and 18)

[66] S. Mangan and U. Alon. Structure and function of the feed-forward loop network motif. PNAS **100**, 11980 (2003). (cited on pages 16, 85, and 86)

[67] A. V. Hill. Proceedings of the physiological society: January 22, 1910. The Journal of Physiology **40**, i (1910). (cited on p. 16)

[68] J. Sneyd, K. Tsaneva-Atanasova, V. Reznikov, Y. Bai, M. J. Sanderson, and D. I. Yule. A method for determining the dependence of calcium oscillations on inositol trisphosphate oscillations. Proc Natl Acad Sci U S A **103**, 1675 (2006). (cited on pages 17, 27, 31, 40, 41, and 117)

[69] S. Basu, R. Mehreja, S. Thiberge, M.-T. Chen, and R. Weiss. Spatiotemporal control of gene expression with pulse-generating networks. Proc Natl Acad Sci U S A **101**, 6355 (2004). (cited on p. 17)

[70] V. Chickarmane and C. Peterson. A computational model for understanding stem cell, trophectoderm and endoderm lineage determination. PLoS One **3**, e3478 (2008). (cited on p. 17)

[71] D. T. Gillespie. *Markov Processes: An Introduction for Physical Scientists* (Academic, 1992). (cited on p. 18)

[72] G. Zlokarnik, P. A. Negulescu, T. E. Knapp, L. Mere, N. Burres, L. Feng, M. Whitney, K. Roemer, and R. Y. Tsien. Quantitation of transcription and clonal selection of single living cells with beta-lactamase as reporter. Science **279**, 84 (1998). (cited on p. 18)

[73] Z. Waks, A. M. Klein, and P. A. Silver. Cell-to-cell variability of alternative rna splicing. Mol Syst Biol **7**, 506 (2011). (cited on p. 18)

[74] C. W. Gardiner. *Handbook of Stochastic Methods* (Springer, 1985). (cited on pages 19 and 77)

[75] D. R. Rigney and W. C. Schieve. Stochastic model of linear, continuous protein synthesis in bacterial populations. J Theor Biol **69**, 761 (1977). (cited on p. 19)

[76] D. R. Rigney. Stochastic model of constitutive protein levels in growing and dividing bacterial cells. J Theor Biol **76**, 453 (1979). (cited on p. 19)

[77] E. Helfand. Numerical integration of stochastic differential equations. The Bell System Technical Journal (1979). (cited on p. 21)

[78] X. Cai and J. Wen. Efficient exact and k-skip methods for stochastic simulation of coupled chemical reactions. J Chem Phys **131**, 064108 (2009). (cited on p. 21)

[79] A. Samant and D. G. Vlachos. Overcoming stiffness in stochastic simulation stemming from partial equilibrium: a multiscale monte carlo algorithm. J Chem Phys **123**, 144114 (2005). (cited on p. 21)

[80] W. E, D. Liu, and E. Vanden-Eijnden. Nested stochastic simulation algorithm for chemical kinetic systems with disparate rates. J Chem Phys **123**, 194107 (2005). (cited on p. 21)

[81] G. Balázsi, A. van Oudenaarden, and J. J. Collins. Cellular decision making and biological noise: from microbes to mammals. Cell **144**, 910 (2011). (cited on p. 21)

[82] M. J. Berridge. Inositol trisphosphate and calcium signaling. Ann N Y Acad Sci **766**, 31 (1995). (cited on pages 23 and 31)

[83] M. J. Berridge. Inositol trisphosphate and calcium signalling mechanisms. Biochim Biophys Acta **1793**, 933 (2009). (cited on pages 23 and 31)

[84] K. P. Hoeflich and M. Ikura. Calmodulin in action: diversity in target recognition and activation mechanisms. Cell **108**, 739 (2002). (cited on p. 23)

[85] F. C. Stevens. Calmodulin: an introduction. Can J Biochem Cell Biol **61**, 906 (1983). (cited on p. 23)

[86] D. E. Clapham. Calcium signaling. Cell **131**, 1047 (2007). (cited on pages 23 and 24)

[87] T. Miyakawa, A. Maeda, T. Yamazawa, K. Hirose, T. Kurosaki, and M. Iino. Encoding of ca2+ signals by differential expression of ip3 receptor subtypes. EMBO J **18**, 1303 (1999). (cited on p. 25)

[88] T. Meyer and L. Stryer. Molecular model for receptor-stimulated calcium spiking. Proc Natl Acad Sci U S A **85**, 5051 (1988). (cited on pages 27, 40, 41, and 117)

[89] A. Goldbeter, G. Dupont, and M. J. Berridge. Minimal model for signal-induced ca2+ oscillations and for their frequency encoding through protein phosphorylation. Proc Natl Acad Sci U S A **87**, 1461 (1990). (cited on pages 27, 40, 41, 43, and 117)

[90] A. Atri, J. Amundson, D. Clapham, and J. Sneyd. A single-pool model for intracellular calcium oscillations and waves in the xenopus laevis oocyte. Biophys J **65**, 1727 (1993). (cited on pages 27, 40, 41, and 117)

[91] Y. X. Li and J. Rinzel. Equations for insp3 receptor-mediated [ca2+]i oscillations derived from a detailed kinetic model: a hodgkin-huxley like formalism. J Theor Biol **166**, 461 (1994). (cited on pages 27, 40, 41, and 117)

[92] J. Sneyd, A. LeBeau, and D. Yule. Traveling waves of calcium in pancreatic acinar cells: model construction and bifurcation analysis. Physica D: Nonlinear Phenomena **145**, 158 (2000). (cited on pages 27, 41, and 117)

[93] T. Matsu-ura, T. Michikawa, T. Inoue, A. Miyawaki, M. Yoshida, and K. Mikoshiba. Cytosolic inositol 1,4,5-trisphosphate dynamics during intracellular calcium oscillations in living cells. J Cell Biol **173**, 755 (2006). (cited on pages 27 and 62)

[94] A. NOVICK and L. SZILARD. Description of the chemostat. Science **112**, 715 (1950). (cited on p. 32)

[95] A. NOVICK and L. SZILARD. Experiments with the chemostat on spontaneous mutations of bacteria. Proc Natl Acad Sci U S A **36**, 708 (1950). (cited on p. 32)

[96] G. M. Whitesides, E. Ostuni, S. Takayama, X. Jiang, and D. E. Ingber. Soft lithography in biology and biochemistry. Annu Rev Biomed Eng **3**, 335 (2001). (cited on p. 32)

[97] W. S. Nguyen, N.T. *Fundamentals and Applications of Microfluidics* (Artech House, 2006). (cited on p. 32)

[98] M. R. Bennett and J. Hasty. Microfluidic devices for measuring gene network dynamics in single cells. Nat Rev Genet **10**, 628 (2009). (cited on p. 32)

[99] S. Mittal, I. Y. Wong, W. M. Deen, and M. Toner. Antibody-functionalized fluid-permeable surfaces for rolling cell capture at high flow rates. Biophys J **102**, 721 (2012). (cited on p. 32)

[100] J. I. V. Mazet, D. Brie. Baseline spectrum estimation using half-quadratic minimization. EUSIPCO 2004 305–308 (2004). (cited on p. 38)

[101] T. O'Haver. Peak finding and measurement. (cited on pages 38 and 46)

[102] M. J. Smit, S. M. Bloemers, R. Leurs, L. G. Tertoolen, A. Bast, S. W. de Laat, and H. Timmerman. Short-term desensitization of the histamine h1 receptor in human hela cells: involvement of protein kinase c dependent and independent pathways. Br J Pharmacol **107**, 448 (1992). (cited on p. 52)

[103] F. Gnad, J. Estrada, and J. Gunawardena. Proteus: a web-based, context-specific modelling tool for molecular networks. Bioinformatics (2012). (cited on pages 60 and 62)

[104] L. Ashall, C. A. Horton, D. E. Nelson, P. Paszek, C. V. Harper, K. Sillitoe, S. Ryan, D. G. Spiller, J. F. Unitt, D. S. Broomhead, D. B. Kell, D. A. Rand, V. Sée, and M. R. H. White. Pulsatile stimulation determines timing and specificity of nf-kappab-dependent transcription. Science **324**, 242 (2009). (cited on p. 65)

[105] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon. Network motifs in the transcriptional regulation network of escherichia coli. Nat Genet **31**, 64 (2002). (cited on p. 65)

[106] N. Nandagopal and M. B. Elowitz. Synthetic biology: integrated gene circuits. Science **333**, 1244 (2011). (cited on pages 66 and 106)

[107] A. S. Khalil, T. K. Lu, C. J. Bashor, C. L. Ramirez, N. C. Pyenson, J. K. Joung, and J. J. Collins. A synthetic biology framework for programming eukaryotic transcription functions. Cell **150**, 647 (2012). (cited on pages 66 and 106)

[108] G. Hornung and N. Barkai. Noise propagation and signaling sensitivity in biological networks: a role for positive feedback. PLoS Comput Biol **4**, e8 (2008). (cited on pages 69, 70, 71, 72, 85, 89, 100, and 105)

[109] P. Hersen, M. N. McClean, L. Mahadevan, and S. Ramanathan. Signal processing by the hog map kinase pathway. Proc Natl Acad Sci U S A **105**, 7165 (2008). (cited on pages 71, 84, and 105)

[110] J. T. Mettetal, D. Muzzey, C. Gómez-Uribe, and A. van Oudenaarden. The frequency dependence of osmo-adaptation in saccharomyces cerevisiae. Science **319**, 482 (2008). (cited on p. 71)

[111] M. R. Bennett, W. L. Pang, N. A. Ostroff, B. L. Baumgartner, S. Nayak, L. S. Tsimring, and J. Hasty. Metabolic gene regulation in a dynamically changing environment. Nature **454**, 1119 (2008). (cited on p. 71)

[112] T. S. Shimizu, Y. Tu, and H. C. Berg. A modular gradient-sensing network for chemotaxis in escherichia coli revealed by responses to time-varying stimuli. Mol Syst Biol **6**, 382 (2010). (cited on p. 71)

[113] C. D. Cox, J. M. McCollum, M. S. Allen, R. D. Dar, and M. L. Simpson. Using noise to probe and characterize gene circuits. Proc Natl Acad Sci U S A **105**, 10809 (2008). (cited on p. 71)

[114] J. Paulsson. Summing up the noise in gene networks. Nature **427**, 415 (2004). (cited on pages 72 and 75)

[115] R. Guantes, J. Estrada, and J. Poyatos. Trade-offs and noise tolerance in signal detection by genetic circuits. Plos One **5** (2010). (cited on pages 72 and 101)

[116] C. Tan, F. Reza, and L. You. Noise-limited frequency signal transmission in gene circuits. Biophys J **93**, 3753 (2007). (cited on p. 73)

[117] S. H. Strogatz. *Nonlinear Dynamics And Chaos: With Applications To Physics, Biology, Chemistry, And Engineering* (Westwiev press, 1994). (cited on p. 73)

[118] J. Keizer. *Statistical Thermodynamics of Nonequilibrium Processes* (Springer-Verlag, 1987). (cited on p. 75)

[119] M. Samoilov, A. Arkin, and J. Ross. Signal processing by simple chemical systems. The Journal of Physical Chemistry A **106**, 10205 (2002). (cited on p. 78)

[120] A. Mugler, B. Grinshpun, R. Franks, and C. H. Wiggins. Statistical method for revealing form-function relations in biological networks. Proc Natl Acad Sci U S A **108**, 446 (2011). (cited on pages 82, 98, and 106)

[121] B. M. C. Martins and P. S. Swain. Trade-offs and constraints in allosteric sensing. PLoS Comput Biol **7**, e1002261 (2011). (cited on pages 82, 98, and 106)

[122] S. Hooshangi, S. Thiberge, and R. Weiss. Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. Proc Natl Acad Sci U S A **102**, 3581 (2005). (cited on pages 84 and 105)

[123] S. Kaplan, A. Bren, A. Zaslaver, E. Dekel, and U. Alon. Diverse two-dimensional input functions control bacterial sugar genes. Mol Cell **29**, 786 (2008). (cited on pages 93 and 106)

[124] S. Kaplan, A. Bren, E. Dekel, and U. Alon. The incoherent feed-forward loop can generate non-monotonic input functions for genes. Mol Syst Biol **4**, 203 (2008). (cited on pages 93, 94, 105, and 106)

[125] E. Richet. Synergistic transcription activation: a dual role for crp in the activation of an escherichia coli promoter depending on malt and crp. EMBO J. **19**, 5222–5232 (2000). (cited on pages 93 and 94)

[126] T. Sohka, R. A. Heins, R. M. Phelan, J. M. Greisler, C. A. Townsend, and M. Ostermeier. An externally tunable bacterial band-pass filter. Proc Natl Acad Sci U S A **106**, 10135 (2009). (cited on pages 93, 94, and 105)

[127] R. J. Prill, P. A. Iglesias, and A. Levchenko. Dynamic properties of network motifs contribute to biological network organization. PLoS Biol **3**, e343 (2005). (cited on p. 96)

[128] S. Tay, J. J. Hughey, T. K. Lee, T. Lipniacki, S. R. Quake, and M. W. Covert. Single-cell nf-kappab dynamics reveal digital activation and analogue information processing. Nature **466**, 267 (2010). (cited on p. 96)

[129] U. Alon. Network motifs: theory and experimental approaches. Nat Rev Genet **8**, 450 (2007). (cited on pages 96 and 97)

[130] J. W. Locasale. Signal duration and the time scale dependence of signal integration in biochemical pathways. BMC Syst Biol **2**, 108 (2008). (cited on p. 96)

[131] M. Behar, N. Hao, H. G. Dohlman, and T. C. Elston. Dose-to-duration encoding and signaling beyond saturation in intracellular signaling networks. PLoS Comput Biol **4**, e1000197 (2008). (cited on p. 96)

[132] S. Mangan, A. Zaslaver, and U. Alon. The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks. J. Mol. Biol. **334**, 197 (2003). (cited on p. 97)

[133] L. Wang, J. Xin, and Q. Nie. A critical quantity for noise attenuation in feedback systems. Plos Computational Biology **6** (2010). (cited on pages 97, 99, 100, and 106)

[134] V. Shahrezaei, J. Ollivier, and P. Swain. Colored extrinsic fluctuations and stochastic gene expression. Molecular systems biology **4** (2008). (cited on p. 97)

[135] S. Ishihara, K. Fujimoto, and T. Shibata. Cross talking of network motifs in gene regulation that generates temporal pulses and spatial stripes. Genes Cells **10**, 1025 (2005). (cited on p. 105)

[136] M. L. Simpson, C. D. Cox, and G. S. Sayler. Frequency domain analysis of noise in autoregulated gene circuits. Proc Natl Acad Sci U S A **100**, 4551 (2003). (cited on p. 105)